DEGRADATION MODELING AND MONITORING OF ENGINEERING SYSTEMS USING FUNCTIONAL DATA ANALYSIS

A Thesis
Presented to
The Academic Faculty

by

Rensheng Zhou

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2012

DEGRADATION MODELING AND MONITORING OF ENGINEERING SYSTEMS USING FUNCTIONAL DATA ANALYSIS

Approved by:

Dr. Nagi Gebraeel, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Nicoleta Serban, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Paul Kvam
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Jianjun Shi
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Vladimir Svetnik
Biometrics Research Department
Merck & Co.

Date Approved: October 30, 2012

To my parents and my sister,

for always being supportive.

ACKNOWLEDGEMENTS

First and foremost, I would like to express the deepest gratitude to my advisors, Prof. Nagi Gebraeel and Prof. Nicoleta Serban. Without their guidance, support and encouragement, it would be impossible for me to complete this thesis all by myself. It is my great honour to be their students and I believe their wisdom and insight will benefit me through my whole life.

I am extremely grateful to my committee members: Dr. Paul Kvam, Dr. Jianjun Shi, and Dr. Vladimir Svetnik (Merck) for serving as my thesis committees. Their insightful suggestions have made this thesis in a much better shape.

I would also like to thank my teammates within the prognostics research group: Linkan Bian, Alaa Elwany, Xiaolei Fang and Li Hao. The discussions with them have always been fruitful. This is an excellent group and I would like to dedicate my appreciation to them.

On the personal level, I would like to express my earnest gratefulness to my friends Ruyun Feng, Xuefeng Gao, Yibiao Lu, Ningyao Zhang and Shipeng Zheng. They are always beside me to share my joy and sadness. It is a great fortune for me to befriended with them. I would also like to extend my appreciations to Jie Gong, Yue Liu, Lei Ma, Shubhankar Ray, Lingyan Ruan, David and Nancy Scott, Mikhail Traskin and Jinjing Wang for their help in various occasions.

Last but not least, I would like to express my love and appreciation to my parents and my sister, who have always supported me. Their love always accompanies me and encourages me to overcome any difficulties in my life. This thesis is dedicated to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

In this thesis, we develop several novel degradation models based on techniques from functional data analysis. These models are suitable for characterizing different types of sensor-based degradation signals, whether they are censored at a certain fixed time point or truncated at the failure threshold. Our proposed models can also be easily extended to accommodate for the effects of environmental conditions on degradation processes. Unlike many existing degradation models that rely on the existence of a historical sample of complete degradation signals, our modeling framework is well-suited for modeling complete as well as incomplete (sparse and fragmented) degradation signals. We utilize these models to predict and continuously update, in real time, the residual life distributions of partially degraded components. We assess and compare the performance of our proposed models and existing benchmark models by using simulated signals and real world data sets. The results indicate that our models can provide a better characterization of the degradation signals and a more accurate prediction of a system's lifetime under different signal scenarios. Another major advantage of our models is their robustness to the model mis-specification, which is especially important for applications with incomplete degradation signals (sparse or fragmented).

CHAPTER I

INTRODUCTION

1.1 Research Background

Most failures of engineering systems result from a gradual and irreversible accumulation of damage that occurs during a system's life cycle. This process is known as *degradation* [12]. In many applications, it can be very difficult to assess and observe physical degradation, especially when real-time observations are required. However, degradation processes are almost always associated with some manifestations that are much easier to observe and monitor over time. Generally, the evolution of these manifestations can be monitored using sensor technology through a process known as *Condition Monitoring* (CM). The observed condition-based signals are known as *degradation signals* [72] and are usually correlated with the underlying physical degradation process. Some examples of degradation signals include vibration signals for monitoring excessive wear in rotating machinery, acoustic emissions for monitoring crack propagation, temperature changes and oil debris for monitoring engine lubrication, and many others.

Degradation modeling characterizes the evolution of degradation signals and predict the lifetime/residual life of systems and/or their components. Within this context, a system's *failure* is typically defined to be the time point when its degradation signal first reaches a critical level known as the *failure threshold*. Most existing degradation models focus on estimating population-related reliability characteristics, such as failure time distributions. In contrast, the overarching objective of this thesis is to predict the residual life distribution (RLD) of a *fielded system*, i.e. a system that is still operating in the field, by using the partial observations of its degradation signal. Another aspect that distinguishes our work from most existing research is that, we focus more on applications with complex degradation signals as well as applications where historical degradation data is sparse. Consequently, we concentrate on developing a nonparametric framework, which raises several theoretical

as well as practical challenges that we address in this thesis.

1.2 Literature Review

One common approach of modeling the degradation signals is to use a parametric model with random/mixed coefficients. In [67], the authors proposed to estimate a time-to-failure distribution by using nonlinear mixed-effects models. A two-stage method was presented for estimating the model parameters. Monte Carlo simulation procedures were deployed for further statistical inferences. In [68], the authors proposed a model with random regression coefficients and a standard-deviation function for analyzing linear degradation data. The parameters within this model were estimated through a likelihood-based approach. In [112], the authors investigated and established asymptotic properties of the ordinary and weighted least squares estimators under the nonlinear mixed-effect model. They used these asymptotic results to obtain point estimates and approximate confidence intervals for percentiles of the failure time distribution. Following a similar approach, the authors of [36] proposed a random effects model for signals with exponential degradation trend. While [67, 68, 112] developed methods to obtain lifetime distributions for a population of components, the authors of [36] focused on computing a RUL distribution for a single operating device using sensory signals. This was accomplished by combining two sources of information: (1) the reliability characteristics of a unit's population; and (2) the real-time sensory signals of the operating unit. Other examples of using mixed coefficients models for degradation modeling include [2, 3, 4, 5, 25, 34, 69, 84, 92, 106, 114].

A stochastic process formulation can be an effective alternative approach to modeling degradation signals. Most research in this area assumes that the error term of degradation path follows a Wiener process or a Gamma process. A major advantage of modeling degradation processes with Wiener processes is that the failure time distribution has a closed-form expression, the inverse Gaussian distribution [15]. The use of Wiener processes and their extensions were investigated in [41, 74, 75, 79, 98, 109, 110]. Gamma processes have been used for modeling monotonically increasing/decreasing degradation signals. The lifetime distribution associated with a Gamma degradation process can be computed in a very

straightforward way due to its monotonic property. Examples of using Gamma processes for modeling degradation signals can be found in [6, 56, 58, 77, 78, 100, 101]. Additionally, some other researchers proposed to use Markovian-based models, assuming that the future degradation state of the unit depends only on the current degradation state. Such examples include [20, 50, 51, 70, 113].

Most existing literature on degradation modeling, including our work in Chapter 2, Chapter 3 and Chapter 4, assume that components are from the same population and are operated under the same environmental conditions. In general, however, components can be operated under different environmental conditions, for instance, different levels of humidity, speeds, loads, temperatures and etc. Environmental conditions can significantly accelerate or decelerate the degradation processes of functioning components. For example, in [35], bearings were run at different rotating speeds and as a result, these bearings degraded at vastly different speeds. One major type of approaches that does take the environmental conditions into account is commonly used for modeling accelerated degradation test (ADT) data. For instance, in [111], the authors proposed a Wiener diffusion process with a time scale transformation that depends upon the level of stress, under which the ADT data was gathered. Similar ideas can be found in [22, 61, 78, 97]. One common characteristic among these approaches is that degradation processes are related to environmental conditions through a simple parametric function. Other examples that incorporate environmental data in degradation models include [46, 49, 60, 90].

1.3 Limitations of Current Methodologies

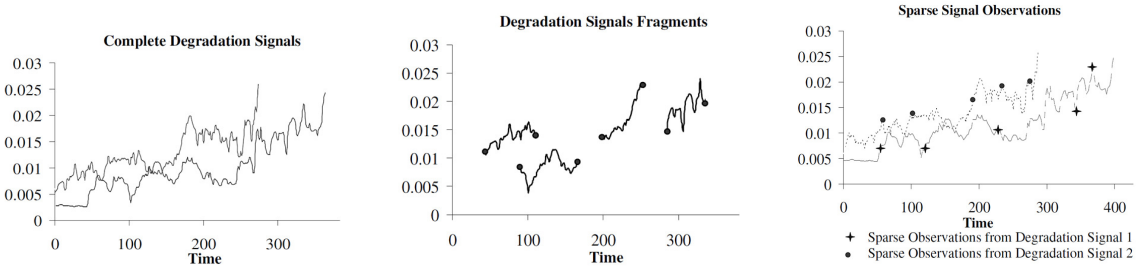
In this section, we discuss the limitations of current methodologies in modeling certain types of degradation signals, which motivate the methodologies presented in the next section.

1.3.1 Modeling *Incomplete* Degradation Signals

Almost all existing degradation models are parametric, i.e., the functional form of the degradation trend is assumed to be known in advance, or identified based on a representative sample of complete degradation signals. A *complete degradation signal* is a continuously observed signal that captures the degradation process of a component from a brand "New

State" to a completely "Failed State"(see Figure 1 (a)).

However, building a database of complete degradation signals can be very expensive and time consuming in applications, such as monitoring of jet engines, turbines, power generating units, structures and bridges, and many others. In these applications, there may be only *incomplete* degradation signals available. In this thesis, we consider two types of incomplete signals, (1) sparse degradation signals (see Figure 1 (c)), or (2) fragmented degradation signals (see Figure 1 (b)). We define fragmented degradation signals as signals with dense observations made over short time intervals. Fragmented signals are important because they can be used to determine the degradation state of a dynamic system, e.g. a power generator. Due to the dynamic nature of such systems, a single sparse observation is not sufficient for characterizing the state of the system. In contrast, the degradation states of relatively static systems, such as crack propagation in bridges and other civil structures, can be adequately characterized by sparse signals.



(a) (b) (c)

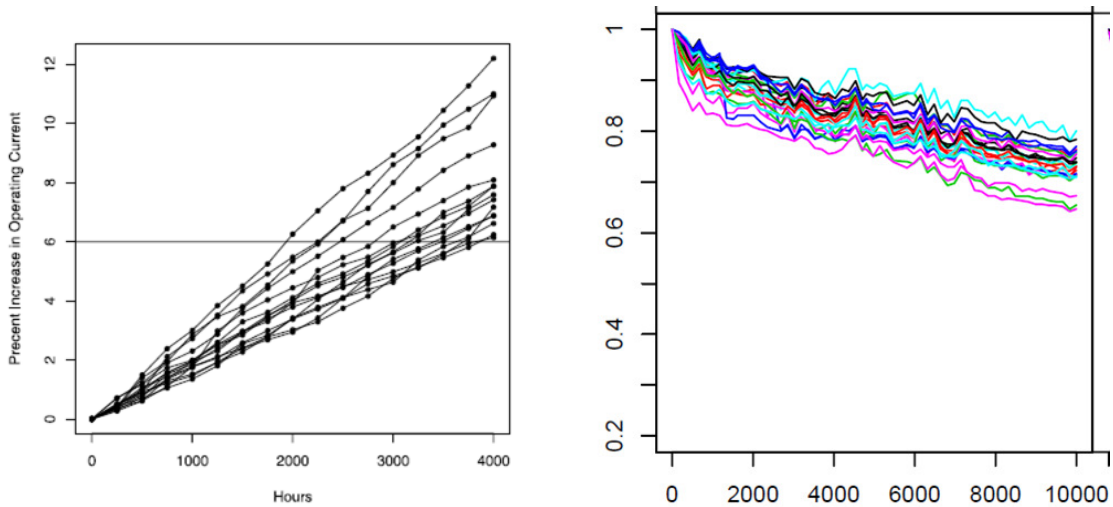
Figure 1: Examples of Complete, Fragemented and Sparse Degradation Signals

One challenge in modeling incomplete degradation signals is that the evolution of the degradation signals cannot be readily assessed to determine the parametric form of the underlying degradation model. This is because one cannot clearly trace how a degradation signal progresses over time from incomplete observations. To overcome this challenge, the underlying degradation model needs to be non-parametric, i.e., the functional form of the degradation trend is assumed to be unknown. Very few work has been done under this assumption. In [88], the authors presented a non-parametric regression model to characterize complete degradation signals at different stress levels and estimated the mean lifetime under

the normal stress. In [107], the authors proposed to model monotonic degradation data using a Gamma process model, in which the shape function was estimated nonparametrically. However, neither [88] nor [107] considered the problem of predicting the lifetime/residual life of a future unit.

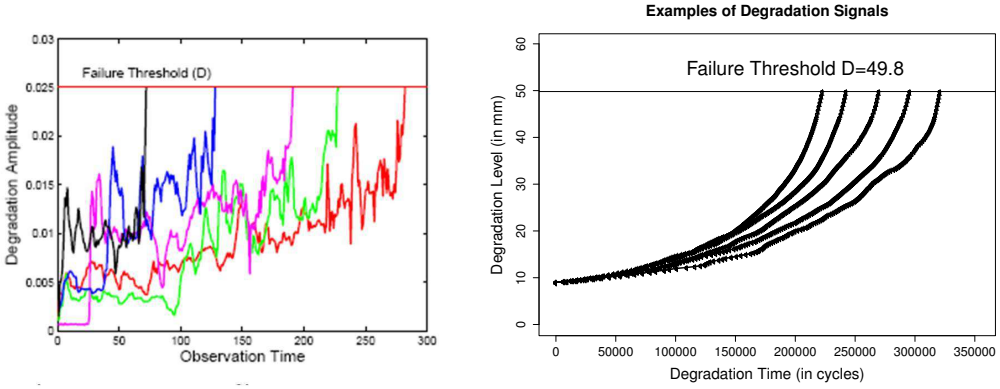
1.3.2 Modeling *Truncated* Degradation Signals

In some engineering applications, it may be possible to continue observing degradation signals after they cross the failure threshold, that is, failures do not necessarily imply system replacements. In this case, practitioners generally observe the degradation signals up to a fixed time point for all systems in the historical data set. In this thesis, degradation signals associated with such applications are referred to as *(time) censored* degradation signals. The LED data in [99] and the laser data in [108] are two examples of censored degradation signals (Figure 2). In contrast, in most other engineering applications, a system is shut down or replaced immediately once its degradation signal reaches the failure threshold, i.e., no further observations can be acquired beyond the failure threshold. We define the degradation signals associated with such applications as *(amplitude) truncated* degradation signals. Examples of truncated degradation signals include the vibration-based degradation signals acquired from a rotating machinery application in [36] and the crack propagation data in [102] (Figure 3).



(a) LED Data (b) Laser Data

Figure 2: Examples of (Time) Censored Degradation Signals



(a) Bearing Degradation Data (b) Crack Propagation Data

Figure 3: Examples of (Amplitude) Truncated Degradation Signals

For truncated degradation signals, the degradation process cannot be recovered beyond the failure threshold due to missing data above the threshold. This challenge has not been widely recognized because most existing degradation models are parametric and these models rely on extrapolation of the degradation process beyond the failure threshold. On the other hand, most existing non-parametric approaches are not directly applicable for modeling truncated degradation signals since they generally assume that the observed random signals share the same time domain. This assumption does not hold for truncated degradation signals. Due to truncation at the failure threshold, truncated degradation signals can only be observed up to their failure time, which varies from one system to another.

1.4 Proposed Methodologies

In this thesis, we focus on predicting the RLD of a fielded system based on the newly observed degradation observations. To this purpose, we follow a two-stage procedure, which is unified within a general Bayesian framework. In the first stage, we develop a degradation model for characterizing the type of degradation signals under investigation. We train the model using historical degradation signals, referred to as *training* degradation signals hereafter. At this stage, we estimate the distribution of the degradation process, which in turn, allows us to recover an approximate (prior) distribution for the residual lifetime according to the failure definition. In the second stage, we predict the RLD of a fielded system by updating the distribution of the degradation process obtained at the first stage

using the partial observations of its degradation signal. This is an online procedure because the update is performed every time a new observation of this fielded system is collected.

In the training stage, we focus on developing non-parametric degradation models based upon tools from functional data analysis (FDA). Different from multivariate data analysis, the focus of FDA is on modeling curves, shapes, images or, more generally, functional observations. An encompassing review of FDA can be found in [80]. FDA techniques have already been applied to many research fields, such as environmental science [26], marketing science [91] and business [105].

One central technique in FDA is Functional Principal Components Analysis (FPCA), which is a common dimension reduction tool, by projecting the original data into new directions, determined by functional principal components, and reducing it to a set of finite dimensional random vectors called functional principal component scores. FPCA was introduced in [81] for growth curves, and was investigated by [8, 21, 40, 73, 83, 89, 115, 116, 117]. In FPCA, the mean and covariance functions are assumed to take unknown functional forms and they are estimated from signal data by following common nonparametric approaches, such as the ones proposed in [27, 80]. The estimated covariance function is then decomposed using the Karhunen-Loève decomposition [47, 66]. The derived eigenfunctions are the so-called functional principal components and they can capture the dominant modes of variations among the given sample of random trajectories. One merit of functional principal components is that they form a compact basis that can approximate the data with as few basis functions as possible.

In Chapter 2, we develop a nonparametric degradation model, based on FPCA, that applies to complete as well as incomplete censored degradation signals. We follow the approach proposed in [117] and estimate the mean and covariance functions using local polynomial smoothers, which is a nonparametric regression approach, based on the pooled data from all units. This enables prediction of individual smooth trajectories even if only one or few measurements are available for a unit. Asymptotic results in [117] demonstrated that this approach is well-suited for applications where only incomplete signals are available for each unit. The work in this chapter is published in [120].

However, preliminary empirical studies in [119] show that traditional FPCA framework is not suitable for modeling truncated degradation signals because the degradation signals beyond the failure threshold are unobservable and therefore missing. Furthermore, the missing data is not at random, thus, inducing bias in the estimation of the mean and covariance functions of the degradation process. Inaccurate estimation of the mean and covariance functions leads to poor performance of FPCA in recovering individual degradation trends and consequently inaccurate residual life prediction. To the best of our knowledge, there are no existing non-parametric regression approaches that can be applied to obtain accurate estimation of the mean and covariance functions based on truncated degradation signals.

To overcome the challenge in modeling truncated degradation signals, we propose two methodologies in Chapter 3 and Chapter 4, respectively. In Chapter 3, we propose a signal transformation procedure, which transforms degradation signals by axis rotation. After transformation, degradation signals that are truncated in the original time domain are no longer truncated with respect to the new amplitude domain. Consequently, we can estimate the mean and covariance functions accurately based on the transformed degradation signals. This allows FPCA to provide an accurate characterization of the degradation trend in the new domain. More details about the signal transformation procedure will be covered in Chapter 3. The work in this chapter is published in [119].

In Chapter 4, we propose another approach for modeling truncated degradation signals by using functional time warping techniques. The primary goal of time warping, also known as curve registration/synchronization, is to align each curve by warping the time domains such that the warped curves maximally coincide with a *template* or *common shape* function [64], or is to optimize pairwise alignments [94, 93]. Various time warping approaches have been proposed to date. Some important examples include landmark registration [10, 54], dynamic time warping [103, 104], functional convex synchronization [64], nonparametric maximum-likelihood warping [38, 85] and self-modeling registration [37, 53, 96]. Other examples inlcude [11, 30, 32, 39, 43, 52, 63]. In this chapter, we pursue a self-modeling registration approach because it allows estimation of the common shape function by borrowing information across signals. The underlying assumption of this method is that systems

degrade according to a common shape function, but at different degradation rates. The degradation signals of the systems are synchronized using a random time warping process, which transforms the common trend function into degradation processes that progress at varying rates. The methodology is partially inspired by [96], which proposed a Bayesian hierarchical curve registration (BHCR) model. Although the original BHCR model was devised only for signals that share the same time domain, we relax this assumption and demonstrate that our modified time warping model is well-suited for modeling truncated signals. The work in this chapter has already been submitted for publication.

As mentioned earlier, our work in the previous chapters assume that all components are operated under the same environments. In contrast, in Chapter 5, we will focus on modeling degradation signals that may have been collected under different types of environmental conditions. In this chapter, we assume that the environmental conditions can be categorized into discrete types, and they are time-invariant. More specifically, we consider two cases, (1) historical degradation signals were collected under known environmental states, and (2) environmental conditions were unknown during the acquisition process for historical data. For the first case, we follow a similar idea in [44] and use a functional linear discriminant analysis algorithm to classify the environment state of the fielded component. Examples of using other methodologies on classifying functional data include functional logistic discriminant analysis in [1, 59], functional support vector machine in [76, 86] and etc. For the second case, an additional step is needed to cluster the historical degradation signals. More specifically, we need to determine the number of environmental types and the cluster membership for all the training components. Generally speaking, there are two types of approaches in functional clustering analysis, (1) hard clustering methods ([14, 17, 31, 87, 95]) that divides a set of random functions into mutually exclusive clusters; and (2) soft clustering methods ([28, 45, 82, 118]) that assume the cluster membership is random and follows a multinomial distribution. In this chapter, we will focus on integrating a soft clustering approach within the proposed degradation modeling framework. We plan to submit the work in this chapter for future publication.

1.5 Thesis Organization

This thesis is organized as follows: Chapter 2 presents an approach to modeling censored degradation signals based on FPCA. We first discuss the development of the degradation modeling framework. An Empirical Bayes approach is introduced for updating the degradation distribution of a partially degraded component. The derivation of the remaining lifetime distribution under the Empirical Bayes approach is also presented. For applications involving incomplete degradation signals, it is important to develop a sampling scheme that ensures accurate estimation of model components. To this purpose, we introduce an experimental design for sampling incomplete degradation signals. The performance results of the methodology are evaluated using real-world as well as simulated degradation signals. In Chapter 3, we introduce a non-parametric degradation modeling framework for truncated degradation signals by using FPCA. A signal transformation procedure is developed, which enables us to use FPCA, in the transformed domain, for modeling truncated degradation signals. The performance of this approach is evaluated using simulated degradation signals and a real-world crack growth data set. Another approach for modeling truncated degradation signals is introduced in Chapter 4. This chapter starts with a description of the general problem and its challenges. The model, which is based on the functional time warping analysis technique, is then applied to the prediction of RLD of fielded systems. The performance of the model is investigated using both simulated degradation signals and real-world bearing degradation signals. Some technical details and additional simulation studies are also provided in the appendix. Chapter 5 extends our previous work to the case when degradation signals are observed under different types of environmental conditions. We incorporate the proposed degradation modeling framework with classification and clustering algorithms that can identify the environmental state of the fielded component. The performance is evaluated by using simulated degradation signals as well as vibration-based degradation signals acquired from a rotating machinery setup. Finally, we outline a future work plan in Chapter 6.

CHAPTER II

MODELING CENSORED DEGRADATION SIGNALS BASED ON FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS

In this chapter, we develop a nonparametric degradation framework, based on FPCA, for modeling censored degradation signals. Specifically, we assume that the degradation process has unknown mean and covariance, which can be estimated through a non-parametric approach using a historical database of degradation signals used to estimate the prior distribution of the degradation process. We also propose an Empirical Bayesian method for predicting the degradation of a partially degraded component or system. To assess the prediction performance, we have applied the non-parametric approach to a real world crack growth data set. This case study demonstrates the non-parametric approach introduced is more accurate in residual life prediction as compared to random effects parametric models which impose constrains on the shape of the trend and the covariance functions.

The remainder of the chapter is organized as follows. Section 2.1 discusses the development of our degradation modeling framework. The Empirical Bayes approach for updating the degradation distribution of a partially degraded component is introduced in Section 2.2. The derivation of the remaining lifetime distribution under the Empirical Bayes approach is presented in Section 2.3. In Section 2.4, we introduce an experimental design for sampling incomplete degradation signals. We discuss performance results of our methodology using real-world and simulated degradation signals in Section 2.5 and 2.6, respectively. Some additional results are deferred to the Appendix A.

2.1 Sensor-based Degradation Modeling

We denote the observed degradation signals $S_i(t_{ij})$, for $j = 1, \ldots, m_i$ (m_i is the number of observation time points for signal i) and $i = 1, \ldots, n$ (n is the number of signals) where $\{t_{ij}\}_{j=1,\ldots,m_i}$ are the observation time points in a bounded time domain $[0, M]$ for signal i. Note that M will always be finite since any industrial application has a finite time of

failure. We model the distribution of the signals $S_i(t)$ by borrowing information across multiple degradation signals. We decompose the degradation signal as

$$S_i(t) = \mu(t) + X_i(t) + \sigma\epsilon_i(t) \tag{1}$$

where $\mu(t)$ is the underlying trend of the degradation process and is assumed to be fixed but unknown, $X_i(t)$ represents the random deviation from the underlying degradation trend. We also assume $X_i(t)$ and $\epsilon_i(t)$ are independent.

The model in (9) is a general decomposition for functional data with various modeling alternatives and assumptions for the model components: $\mu(t)$, $X_i(t)$, and $\epsilon_i(t)$. In this chapter, we discuss one such modeling alternative which applies to sparse and fragmented signals as well as to complete signals and it applies under the assumption that the observation time points $\{t_{ij}\}_{j=1,\ldots,m_i}$ are fixed but not necessarily equally spaced and the assumption that the error terms $\epsilon_i(t)$ are independent and identically distributed. Deviations from these assumptions may require some modifications to the modeling approach discussed in this chapter.

In our modeling approach, the degradation signal $S_i(t)$ follows a stochastic process with mean $\mu(t)$ and stochastic deviations $X_i(t)$ with mean zero and covariance $\text{cov}(t, t')$. The mean function $\mu(t)$ and the covariance surface $\text{cov}(t, t')$ are both assumed to be non-parametric, i.e., no pre-specified assumption on their shape. This generalized assumption encompasses the particular cases developed earlier in [33, 36], which assume a linear trend, $\mu(t) = \alpha + \beta t$ where $\alpha \sim N(0, \delta_\alpha)$ and $\beta \sim N(0, \delta_\beta)$, and parametric covariance structure $\text{cov}(t, t') = \delta_\alpha + \delta_\beta tt'$.

The following steps discuss how we estimate the mean function and the covariance surface of our degradation model.

Step 1: We use nonparametric methods to estimate the mean $\mu(t)$. In this chapter, we use local quadratic smoothing [27] to allow estimation of the mean function under general settings including complete and incomplete (sparse and fragmented) signals. The bandwidth parameter, which controls the smoothing level, is selected using the leave-one-curve-out cross-validation method [83]. Alternative estimation methods include decomposition of the

mean function using a basis of functions (e.g. splines, Fourier, wavelets) and estimate the coefficients using parametric methods. These alternative methods will apply under various signal behaviors (e.g. smooth vs. with sharp changes, uniformly vs. non-uniformly sampled).

Step 2: The covariance surface is estimated using the demeaned data, $S_i(t) - \hat{\mu}(t)$, where $\hat{\mu}(t)$ is the local quadratic smoothing estimate of $\mu(t)$. Using the Karhunen-Loéve decomposition [47, 66], the covariance, $\text{cov}(t, t') = Cov(S_i(t), S_i(t'))$ can be expressed as follows:

$$\text{cov}(t, t') = \sum_{k=1}^{\infty} \lambda_k \phi_k(t) \phi_k(t'), \ t, t' \in [0, M], \tag{2}$$

where $\phi_k(t)$ for $k = 1, 2, \ldots$ are the associated eigenfunctions with support $[0, M]$ and $\lambda_1 \geq \lambda_2 \geq \ldots$ are the ordered eigenvalues. Based on this decomposition, the deviations from the underlying degradation trend $X_i(t)$ are decomposed using the following expression:

$$X_i(t_{ij}) = \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t_{ij}) \tag{3}$$

where ξ_{ik} called *scores* are uncorrelated random effects with mean zero and variance $\mathbb{E}(\xi_{ik}^2) = \lambda_k$. The decomposition in equation (10) is an infinite sum. Generally, only a small number of eigenvalues are commonly significantly non-zero. For the eigenvalues which are approximately zero, the corresponding scores will also be approximately zero. Consequently, we use a truncated version of this decomposition. Therefore, expression (10) can be approximated as follows:

$$X_i(t_{ij}) = \sum_{k=1}^{K} \xi_{ik} \phi_k(t_{ij}) \tag{4}$$

where K is the number of significantly non-zero eigenvalues. We select K to minimize the modified Akaike criterion defined by [117].

In the statistical literature, this method has been coined Functional Principal Component Analysis (FPCA). The key reference for FPCA is [80]. Another important reference is [117], in which the authors derived theoretical results for model parameter consistency and asymptotic (n large) distribution results under the assumption that the scores follow a normal distribution.

An alternative method for estimating the covariance function of the process $X_i(t)$ is decomposing the covariance function as in equation (2) where the basis of functions $\{\phi_k,\ k = 1, 2, \ldots\}$ is fixed [42] . However, this approach doesn't allow dimensionality reduction in the same way FPCA does and it is not theoretically founded.

2.2 Degradation Model Updating

Next, we consider a component operating in the field called *fielded component*. Assume that we have observed its degradation signal at a vector of time $\mathbf{t} = (t_1, \ldots, t_{m^*})$; therefore, $S(\mathbf{t})$ denotes the observed signal of the testing component, m^* represents the number of observations and $t^* = t_{m^*}$ denotes the latest observation time. In this section, we introduce an Empirical Bayes approach which allows real-time updating of the distribution of the degradation process for partially degraded components given the observed signal $S(\mathbf{t})$ and the prior distribution of the scores ξ_{ik} for $k = 1, 2, \ldots$. The prior distribution of the scores is estimated empirically from a set of historical degradation signals.

Proposition 1 illustrates the updating procedure assuming that the prior distribution of the scores is normal and assuming that the mean function $\mu(t)$ and the basis of functions $\phi_k(t), k = 1, \ldots, K$ are fixed. The proof of this proposition follows from the theory of Bayesian linear models.

Proposition 1. *Assuming that $S(t)$ follows*

$$S(t) = \mu(t) + \sum_{k=1}^{K} \xi_k \phi_k(t) + \epsilon(t)$$

where the prior distribution of ξ_k is $N(0, \lambda_k)$ with $\xi_1, \ldots, \xi_K$ uncorrelated; $\epsilon(t)$ are independent of ξ_k for $k = 1, \ldots, K$; the distribution of $\epsilon(t)$ is $N(0, \sigma^2)$ with σ^2 fixed. It follows that the posterior distribution of the scores is

$$(\xi_1^*, \ldots, \xi_K^*)' \sim N(Cd, C) \text{ where}$$

$$C = \left(\frac{1}{\sigma^2} P(\mathbf{t})' P(\mathbf{t}) + \Lambda^{-1}\right)^{-1} \text{ and } d = \frac{1}{\sigma^2} P(\mathbf{t})' (S(\mathbf{t}) - \mu(\mathbf{t}))$$

with

$$S(\mathbf{t}) = (S(t_1), \ldots, S(t_{m^*}))', \mu(\mathbf{t}) = (\mu(t_1), \ldots, \mu(t_{m^*}))'$$
$$\Lambda = diag(\lambda_1, \ldots, \lambda_K), P(\mathbf{t}) = \begin{pmatrix} \phi_1(t_1) & \ldots & \phi_K(t_1) \\ \ldots & \ldots & \ldots \\ \phi_1(t_{m^*}) & \ldots & \phi_K(t_{m^*}) \end{pmatrix} \tag{5}$$

In Proposition 1, the prior distribution of the scores is specified by the variance parameters $\lambda_k,\ k = 1, \ldots, K$, which are estimated using the degradation model in Section 2.1 and based on a set of incomplete or complete training degradation signals. Specifically, we first apply Functional Principal Component Analysis on the historical degradation signals which will further provide estimates for the variance parameters $\lambda_k,\ k = 1, \ldots, K$ and the eigenfunctions $\phi_k,\ k = 1, \ldots, K$. Based on these estimates, we obtain the posterior distributions of the updated scores $\xi_1^*, \ldots, \xi_K^*$ since the matrix C and the vector d are fully determined by the eigenvalues $\lambda_k,\ k = 1, \ldots, K$ and the eigenfunctions $\phi_k,\ k = 1, \ldots, K$. The expectation of the posterior scores is non-zero and therefore we denote the posterior mean function $\mu^*(t) = \mu(t) + \sum_{k=1}^{K} E(\xi_k^*)\phi_k(t)$.

Following Proposition 1, the expectation of the posterior distribution follows the same formula as the conditional expectation estimator in [117] equation (4). Generally, this similarity applies under the empirical Bayesian prior derived from FPCA. On the other hand, the sampling distribution of the conditional expectation estimator in [117] is different from the posterior distribution of $\xi_k^*, k = 1, \ldots, K$ since their variances are not equal. Moreover, the conditional expectation estimator and its mean estimation error in [117] is conditional on the training observations whereas the posterior distribution in Proposition 1 is conditional on the observations of a new component.

The advantage of this Bayesian framework is that it unifies the conditional expectation estimation and prediction into a procedure which allows updating the distribution of the degradation process for a new component. We can therefore use the posterior distribution of the scores for a partially degraded signal to estimate the distribution of various statistical summaries including the lifetime at a specified degradation level and estimation of the degradation level at a specified time. In the next section, we discuss one specific application

to this updating framework: residual life estimation.

2.3 Remaining Life Distribution

In this chapter, we focus our attention on engineering applications where a soft-failure of a system occurs once its underlying degradation process reaches a predetermined critical threshold. This critical threshold is commonly used to initiate maintenance activities such as repair and/or component replacement well in advance of catastrophic failure. Consequently, degradation data can still be observed beyond the critical threshold. In this section, we describe how our degradation modeling framework is applied to estimate the distribution of remaining life up to a degradation threshold of partially degraded systems.

In the remaining of this section, $S^*(\cdot)$ will denote the underlying degradation process of a partially degraded system. Based on the degradation process $S^*(\cdot)$, the failure time of a system is defined as:

$$T = \inf_{t \in [0,M]} \{S^*(t) \geq D\}. \tag{6}$$

One has to bear in mind that T may not exist if the threshold D is set too high; i.e. the component may fail before its degradation signal reaches the threshold. The selection of the failure threshold D is an important problem, however, this aspect is beyond the scope of this chapter. In this work, we assume that T exists, and the threshold D is known a priori. This is a reasonable assumption because in many industrial applications failure/alarm thresholds are usually based on subjective engineering judgement or well-accepted standards, such as the International Standards Organization (ISO). (For example, the ISO 2372 is used for defining acceptable vibration threshold levels for different machine classifications.) A second assumption is that the failure time T is smaller than a maximum failure time M. This assumption is also reasonable as in practice a component may be replaced after a given period of time even if it didn't fail.

The distribution of the residual life (RLD) of a partially degraded component at a fixed time $t^* \in [0, M]$ is estimated assuming that the degradation process $S^*(\cdot)$ of the component follows a posterior distribution based on Proposition 1. We estimate the distribution of the

residual life (RLD) using

$$R(y|t^*) = P(T - t^* \leq y \mid S^*(t) \sim \text{Gaussian}(\mu^*(t), \text{Cov}^*(t,t')),\ t^* \leq T \leq M)$$

where $\mu^*(t)$ and $\text{Cov}^*(t,t')$ are the posterior mean and covariance functions of the degradation process $S^*(\cdot)$. The derivations of $\mu^*(t)$ and $\text{Cov}^*(t,t')$ are based on the results of Proposition 1. We note here that the distribution of the RLD above is not conditional on the observed signal of the partially degraded component but on the posterior distribution of its degradation process; since the degradation is only partially observed and most often sparsely sampled, conditioning on the posterior distribution will generally provide a more accurate RLD estimator since we incorporate the additional information in the training degradation signals.

Furthermore, we estimate RLD under two assumptions:

A.1 The new component has not failed up to the last observation time point t^*; that is, the failure time becomes

$$T = \inf_{t \in [0,M]} \{S^*(t) \geq D\} = \inf_{t \in [t^*,M]} \{S^*(t) \geq D\} := T^*.$$

A.2 We assume the probability that the degradation process $S^*(t)$ crosses back the threshold D after the failure time T^* is negligible, i.e., $P(S^*(T^* + y) < D) \approx 0$ for all $y > 0$. This implies, if we condition on $y \geq T^* - t^* > 0$, which is the same as conditioning on $T^* \leq t^* + y$, $P(S^*(t^* + y) < D | T^* \leq t^* + y) \approx 0$. This further implies

$$P(S^*(t^* + y) \geq D) = P(S^*(t^* + y) \geq D | T^* \leq t^* + y) P(T^* \leq t^* + y) \approx P(T^* \leq t^* + y).$$

Under these two assumptions, the RLD becomes

$$R(y|t^*) = (\text{by A.1})\ P(T^* - t^* \leq y |\ S^*(t)) \approx (\text{by A.2})\ P(S^*(t^* + y) \geq D |\ S^*(t)).$$

The approximation in assumption (A.2) is similar to the approximation in the paper by [67] which assumes that the probability of a negative random slope in the linear model is negligible. One particular case for the assumption A.2 to hold is that the signal is

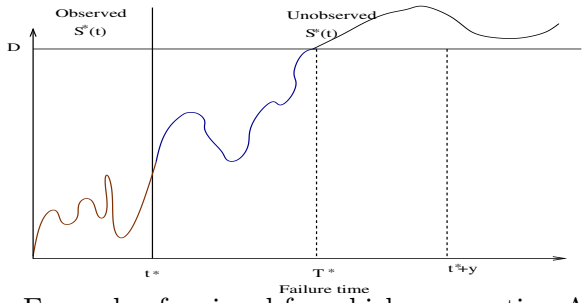


Figure 4: Example of a signal for which assumption A.2 holds.

monotone. However, monotonicity is not a necessary condition. Assumption A.2 also holds for non-monotone signals - an example of such signal is in Figure 4.

Proposition 2 below describes the updating procedure for RLD of a new component given the posterior distribution of its degradation process $S^*(\cdot)$ updated up to time t^*. The proof follows directly as a consequence of Proposition 1.

Proposition 2. *For a new partially degraded component with its degradation process $S^*(\cdot)$ updated up to time t^*, the residual life distribution is given as follows*

$$P(T - t^* \leq y \mid S^*(\cdot), T \geq t^*) = \frac{\Phi_Z\left(g^*(y|t^*)\right) - \Phi_Z\left(g^*(0|t^*)\right)}{1 - \Phi_Z\left(g^*(0|t^*)\right)} \tag{7}$$

where Φ_Z represents the standard normal cumulative distribution function and $g^*(y|t^*) = \frac{\mu^*(t^*+y)-D}{\sqrt{V^*(t^*+y)}}$ with:

$$\mu^*(t^* + y) = \mu(t^* + y) + (Cd)'p(t^* + y)$$

$$V^*(t^* + y) = \textstyle\sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \left[C_{k1,k2}\phi_{k_1}(t^* + y)\phi_{k_2}(t^* + y)\right].$$

In the above equations, $p(t^* + y) = (\phi_1(t^* + y), \ldots, \phi_K(t^* + y))'$, and $C_{k1,k2}$ refers to the (k_1, k_2) element of the matrix C.

One advantage of obtaining the distribution rather than simply a point estimate is that we can also derive a *confidence interval* for the remaining lifetime up to a degradation threshold D. Following the derivation in Proposition 2, a $1-\alpha$ confidence interval for RLD is $[L, U]$ such that

$$P(L \leq T - t^* \leq U|\ S^*(\cdot),\ T \geq t^*) = 1 - \alpha.$$

Since we have one equation with two unknowns, the lower - L and the upper - U tails are

commonly equally weighted, and therefore,

$$\frac{\Phi_Z\left(g^*(U|t^*)\right) - \Phi_Z\left(g^*(0|t^*)\right)}{1 - \Phi_Z\left(g^*(0|t^*)\right)} = 1 - \frac{\alpha}{2} \text{ and } \frac{\Phi_Z\left(g^*(L|t^*)\right) - \Phi_Z\left(g^*(0|t^*)\right)}{1 - \Phi_Z\left(g^*(0|t^*)\right)} = \frac{\alpha}{2}.$$

However, we cannot obtain exact solutions for L and U because we do not have a closed-form expression for the inverse of the cumulative density function of $T - t^*$. For example, the first relationship is equivalent to finding U from $g^*(U|t^*) = z_{\alpha_1}$ where z_{α_1} is the $1 - \alpha_1$ quantile of the normal distribution. Using this equation, we would like to obtain U such that

$$\frac{\mu^*(t^* + U) - D}{\sqrt{V^*(t^* + U)}} = z_{\alpha_1}$$

which is a non-linear function of U and its solution does not have a close form expression. We therefore resort to *parametric bootstrap* [18, 23] to sample from the distribution of $T - t^*$ which will give us a set of realizations from this distribution - $T_1, T_2, \ldots, T_B$. Using these realizations from the distribution of $T - t^*$, we estimate a quantile bootstrap confidence interval.

The confidence interval estimation procedure is as follows. For $b = 1, \ldots, B$,

1. Sample $\xi^b = (\xi_1^b, \ldots, \xi_K^b)$ from the multivariate normal distribution of the posterior scores provided in Proposition 1.

2. Obtain a simulated signal

$$S_b(t) = \mu(t) + \sum_{k=1}^{K} \xi_k^b \phi_k(t)$$

   where $\xi_k^b,\ k = 1, \ldots, K$ are the scores sampled at Step 1.

3. Take $T_b = \inf_{t \in [0,M]}\{S_b(t) \geq D\}$

Using the sampled values $T_1, T_2, \ldots, T_B$, we compute the empirical $\alpha/2$ and $(1 - \alpha/2)$ quantiles, $T_{\alpha/2}$ and $T_{1-\alpha/2}$, respectively. We estimate the upper and lower bound of the confidence interval by $\hat{L} = T_{\alpha/2}$ and $\hat{U} = T_{1-\alpha/2}$. It follows that $[\hat{L}, \hat{U}]$ is an approximate $1 - \alpha$ quantile bootstrap confidence interval for the residual life time of the fielded component.

An additional approach to the (parametric) bootstrap method described above is to (re)sample the signal data resulting in multiple bootstrap samples. For each bootstrap

sample, estimate the residual lifetime using the approach discussed in this chapter; therefore, we obtain a set of realization from the distribution of $T - t^*$. In contrast to the bootstrap method described above, this alternative bootstrap approach requires estimating the FPCA model for each bootstrap sample which is computationally expensive.

2.4 Sampling Scheme

The non-parametric degradation modeling framework introduced in this chapter applies to both complete as well as incomplete degradation signals. For applications involving incomplete degradation signals, it is important to develop a sampling plan that ensures accurate estimation of the mean function and the covariance surface. [117] provides theoretical results on the estimation of the covariance surface using FPCA under *large n but small m_i* for $i = 1, \ldots, n$. In other words, for these results to hold, the observation time points $\{t_{ij}\}_{j=1,\ldots,m_i, i=1,\ldots,n}$ need to cover the time domain, $[0, M]$, densely.

Using the traditional uniform sampling technique, the number of observations per time interval decreases as more signals fail leading to an unbalanced number of observations per time interval - more observations at the beginning of the observation time domain but fewer observations at the end of the time domain. Further, this unbalanced design will result in decreasing estimation accuracy (higher variances) of the mean and covariance estimates at later time points. In order to balance the number of observations per time interval throughout the time domain $[0, M]$, we propose an experimental design using non-uniform sampling. The proposed technique ensures relatively dense coverage of the sampling time domain, $[0, M]$, where M represents the last observation time of the longest possible degradation signal for a given application.

We note here that the sampling technique requires input of M at the beginning of the experiment although M is unknown. It is often the case that in practice, an experimenter will set a timeline at the beginning of the experiment which will specify a limit of how long the experiment will be run (e.g. one year vs one month). This upper limit will specify M. Generally, starting with a lower initial value for M will allow the experimenter to sample densely enough while having the option to update the sampling technique (update M) if

not all training signals have reached the failure threshold by the initial value for M.

The following steps outline a sampling procedure for obtaining sparsely observed and fragmented degradation signals.

Step 1: We begin by performing non-uniform sampling of the time domain $[0, M]$, thus, obtaining a sequence of time points, $0 = t_1 < t_2 < \ldots < t_{m-1} < t_m = M$, for large m. Since only a few components will survive up to the maximum time point, M, we increase the sampling frequency at later time points in order to cover the sampling time domain at the extreme point, M. Consequently, we sample exponentially, i.e. the time interval between two consecutive sampling time points decreases exponentially over time (the decreasing rate is implicitly determined by the value of M and the number of sampling time points).

Step 2: This step provides a potential sampling timetable (or monitoring/observation schedule) for sparsely observed and fragmented degradation signals. We begin by selecting n components. For each component, we select its sampling time points from the set $t_1, \ldots, t_m$ *without any prior knowledge about their degradation process and lifetime.* Next we define two settings.

Setting 1: This setting is used to obtain *sparsely observed degradation signals.* For component i, we randomly sample m_i time points from the set of total time points $\{t_1, \ldots, t_m\}$. This results in a the set of sparse *sampling time points*, $\{t_{i1}, \ldots, t_{im_i}\}$ for this component.

Setting 2: This setting is used to obtain *fragmented degradation signals.* Recall that fragmented signals are obtained by continuously monitoring a component over a short time interval, hence the term "fragment". For component i, we select two or more time points $B_1, B_2, \ldots$ from the set of total time points $\{t_1, \ldots, t_m\}$. These points represent the beginning times of the signal fragments or sampling intervals. The duration of the sampling interval will depend on the type of application, the availability of monitoring/testing equipment, and the associated costs/economics. Consequently, the end time points, $E_1, E_2, \ldots$, will vary from one experiment to another. In other words, for component i, we may have two or more time intervals: $[B_{i,1}, E_{i,1}], [B_{i,2}, E_{i,2}] \ldots$.

Step 3: Finally, we observe the degradation signal for the selected components at the selected time points according to the type of incomplete signals - sparse or fragmented, and

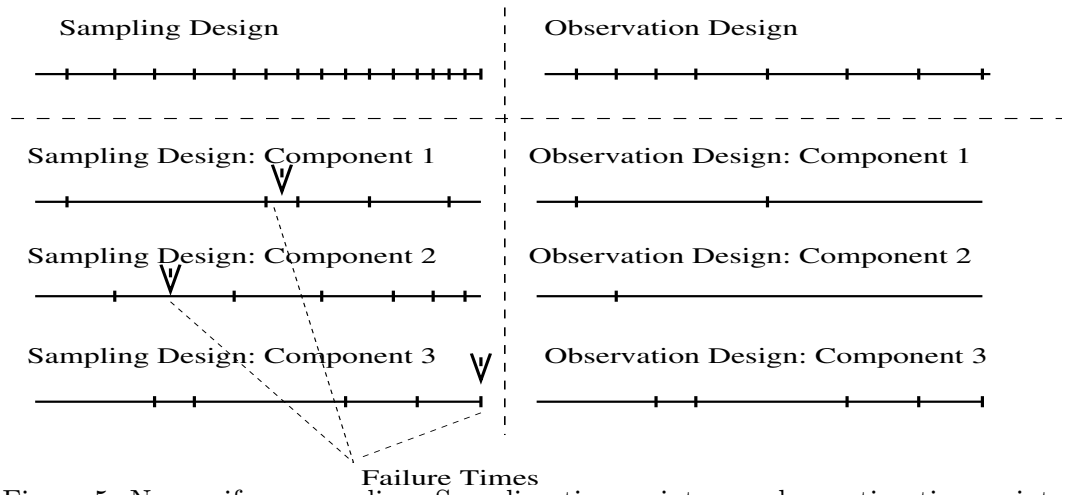


Figure 5: Non-uniform sampling: Sampling time points vs. observation time points.

obtain the set of sampled signals $S_i(t_{ij})$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$.

It is important to stress that we select the sampling time points in Step 2 before observing the degradation signals. Since we do not observe the failure time before selecting the time points, we cannot ensure that the degradation signal will be observed for all selected sampling time points. This is because some components may fail before the latest selected time point. Consequently, the *observation time points* are a subset of the sampling time points and will be less densely sampled close to M since the missing observations (the difference set between sampling and observation time points) will increase in density closer to the upper bound M. In Figure 5, we compare the sampling time points selected at Step 2 to the observation time points for three components. In this example, the sampling time points are non-uniformly selected whereas the observation time points are approximately uniform since for the first two components, we do not observe at the latest times - only the third component fails after its latest sampling time.

Two parameters that are used for tuning the sampling plan are: the total number of sampling time points $m_{total} = \sum_{i=1}^{n} m_i$ and the total number of components n. The more sparsely the signals are observed (m_{total} is small), the more signals we need to observe (n needs to be large). Selecting n and m_{total} optimally is important to ensure accurate modeling of the degradation process at a feasible cost. The larger the number of components n and/or the larger the number of time points m_{total} are the higher the costs associated

with monitoring and testing. Note that selection of n and m_{total} will vary according to the type of application.

2.5 Case Study: Crack Growth Data

In this section, we study crack growth data that can be found in various domains of engineering applications, such as infrastructure (bridges, steel structures), maritime (hulls of oil tankers), aeronautical (aircraft fuselage), energy (vanes of gas turbines), etc. We consider a situation in which crack growth data can be observed from identical units (say several ship hulls, or turbines) up to a predetermined time period, denoted by M in this chapter. A constant threshold, D, is a critical crack length representing a soft failure when maintenance and repair should be performed. Within this context, we assume that catastrophic failure, i.e. hard failure, may occur at a relatively larger crack length.

The data set used in our case study was first published in [102], and has been previously analyzed in other journal articles [16, 55] and the references therein. The specimens in the test were 2.54-mm-thick and 152.4-mm-wide center cracked sheets of 2024-T3 aluminum. The crack propagation signals of these specimens were recorded under identical experimental conditions. In this data set, the crack length was measured in millimeters and the observation time was measured by the cumulative load cycles. More details about this data set can be found in [102]. In this study, we set the soft failure threshold to $D = 27$mm. We provide additional results for another soft threshold in the Appendix [120]. To be consistent with the methodology in this study, the observations are censored at common value $M = 230000$ cycles. A representative example of sparsely sampled degradation signals is in Figure 6(a).

2.5.1 Results and Analysis

We report the prediction accuracy of the remaining life for varying time points t^* defining the latest observation time of a partially degraded component. We consider the following degradation percentiles: 10% (the signal has been observed up to time t^*, which equals to 10% of the lifetime), 20%, ..., 80% and 90%. For each crack, we predict the updated residual lifetime at each of the nine percentiles using the degradation signal observed up to

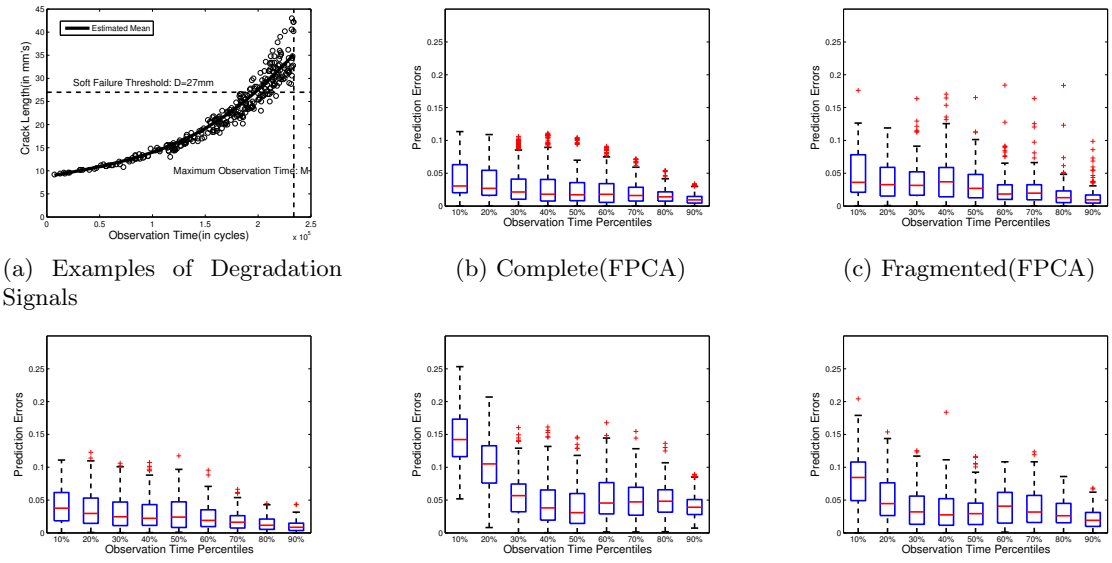
that respective percentile. The number of signals in this study is 59. We randomly select 50 of the total signals as training signals for estimating the model components, and the rest are validation signals for evaluating the performance of our model in predicting residual life. For each validation signal, we use the following error criteria to assess the prediction accuracy:

$$\text{error} = \frac{|\text{Estimated Life - Actual Life}|}{\text{Actual Life}} \tag{8}$$

We replicate the above procedure for 100 times, and report the distribution of the errors across the 100 simulations using a set of boxplots, each boxplot corresponding to a degradation percentile for the testing components and providing the absolute prediction errors for that percentile.

We first discuss the performance of our non-parametric model for complete, sparse and fragmented degradation signals. In each complete degradation signal, we have about 50 observations per signal. To obtain a sparsely observed degradation signal, we randomly sample $m = 6$ observations from each complete signal. We use two intervals per signal to obtain fragmented degradation signals. The results are illustrated in Figure 6(b-d). The results indicate that our nonparametric model performs well for complete as well as incomplete degradation signals, and the performance is better when the incomplete degradation signals are sparse rather than fragmented. Although we have only approximately 10% observations of complete degradation signals under the sparse sampling scenario, the prediction errors do not increase significantly. This observation is important in practice; under budget limitations, one may resort to sparse or fragmented degradation signals without significant loss of predictive capability.

We also demonstrate the benefits of our proposed non-parametric degradation model by comparing it with parametric models as benchmarks. Since the degradation signals have a non-linear trend with a curvature similar to the exponential function, we transform the degradation signals using the natural logarithm in order to linearize the trend and then apply a linear random effects model (henceforth, denoted by 'Log-linear'). Since under the log-transform model, the residual life predictions are inaccurate compared to the non-parametric approach, we consider a double logarithm transformation of the degradation



(a) Examples of Degradation Signals

(b) Complete(FPCA)

(c) Fragmented(FPCA)

(d) Sparse(FPCA)

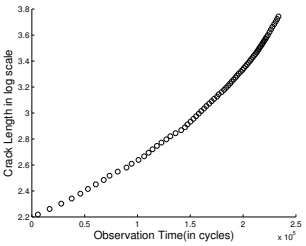
(e) Sparse: Log-linear

(f) Sparse: Log-log-linear

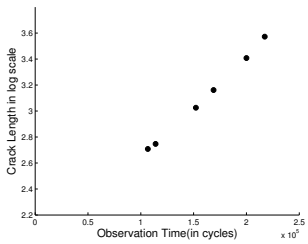
Figure 6: The prediction error of residual life prediction for the crack growth data set.

data (henceforth, denoted by 'Log-log-linear'). The results of the sparse scenario using the parametric models 'Log-linear' and 'Log-log-linear' are reported in Figure 6(e-f), respectively. We find that both parametric models provide less accurate predictions of the residual life than our non-parametric model. This is due to the inaccuracy of the parametric models in capturing the crack propagation trend.

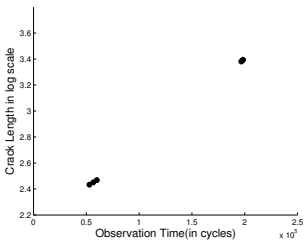
We provide one example in Figure 18 to illustrate the source of the bias of the 'Log-linear' model. In this Figure, the x-axis represents the degradation time and the y-axis represents the crack length, but in the log scale. We have one complete, sparse and fragmented degradation signal in Figure 18(a-c), respectively. If the 'Log-linear' model is the true underlying parametric model, we should see a linear trend in all three plots. This seems to be true in the sparse or fragmented cases (see Figure 18(b-c)). However, for Figure 18(a) showing a complete signal, we note that the degradation trend is still nonlinear; the log-transformation does not linearize the signal (the same applies for the 'log-log' transformation). Therefore, the 'log-linear' model does not accurately capture the crack propagation trend throughout the unit's lifetime. This example shows the potential difficulty of identifying a reasonable parametric model for sparse and fragmented degradation signals and in turn, demonstrates the robustness of our proposed non-parametric model to model mis-specification.



(a) An Example of Complete Signal(in log-scale)



(b) An Example of Sparse Signal(in log-scale)



(c) An Example of Fragmented Signal(in log-scale)

Figure 7: Examples of the crack data under the log scale.

2.6 Simulation Study

In this section, we simulate nonlinear degradation signals from three different models to demonstrate the benefits of using our proposed non-parametric degradation modeling approach. We evaluate our approach in terms of the prediction accuracy of estimating the residual life for complete, sparse and fragmented degradation signals, contrast uniform and non-uniform sampling procedures for acquiring the ensembles of incomplete degradation signals, and also investigate the robustness of our model to violations of its model assumptions.

2.6.1 Simulation Models

The degradation signals are simulated from three different models, and all of them are special cases of the general model (9). More specifically,

- In model 1, we choose $\mu(t) = 30t^2$, $X_i(t) = \xi_1\phi_1(t)$, where $\xi_1 \sim N(0, \frac{45}{4})$, $\phi_1(t) = \sqrt{5}t^2$, $0 \leq t \leq 1$, and $\sigma = 1$.

- In model 2, we choose $\mu(t) = 30t^2$, $X_i(t) = \xi_1\phi_1(t) + \xi_2\phi_2(t)$, where $\xi_1 \sim N(0, 3^2)$, $\phi_1(t) = 2t$, and $\xi_2 \sim N(0, (\frac{3}{2})^2)$, $\phi_2(t) = \sqrt{80}t^2 - \frac{3}{4}\sqrt{80}t$, $0 \leq t \leq 1$.(The coefficients of the eigenfunctions are chosen so that they form an orthonormal functional basis for $0 \leq t \leq 1$).

- In model 3, we choose $\mu(t) = 30t^2 - 2sin(4\pi t)$, $X_i(t) = \xi_1\phi_1(t) + \xi_2\phi_2(t)$, where $\xi_1 \sim N(0, 3^2)$, $\phi_1(t) = 2t$, and $\xi_2 \sim N(0, (\frac{3}{2})^2)$, $\phi_2(t) = \sqrt{80}t^2 - \frac{3}{4}\sqrt{80}t$, $0 \leq t \leq 1$.

We simulate from model 1 because its residual life distribution can be easily derived from training signals and updated using validation signals using the procedure in [36]. The derived residual life distribution can then be utilized as a benchmark to assess the performance of our non-parametric approach.

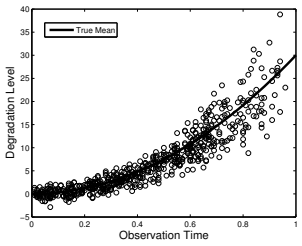
Across all the models, the failure threshold is set to $D = 10$. We generate $n = 100$ "training" signals and $n = 100$ "validation" signals from each model. For a complete signal, we have 51 observations made at an equally spaced grid $c_0, ..., c_{50}$ on $[0, 1]$ with $c_0 = 0, c_{50} = 1$. A sparse or fragmented signal is then sampled from a complete signal such that we observe about 6 observations per signal. The stopping time for each training signal (the last point at which a signal is observed) is generated from Uniform distribution (Uniform$(0.7, 1)$) - our simulation results are insensitive to the selection of the stopping time distribution.

We run simulations for 100 times. For each simulation, we compute the prediction errors at the following degradation percentiles 10%, 20%, ..., 70%, 80% and 90% of the simulated degradation signals.

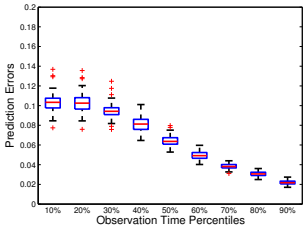
2.6.1.1 Results and Analysis of Model 1

In Figure 8(b-d), we present the boxplots of the prediction errors when using the non-parametric degradation model in this chapter for complete, fragmented and sparse degradation signals. For the sparse scenario, we compare the prediction accuracy of using true parametric model (see Figure 8(e)) and our non-parametric model when signals are uniformly sampled (see Figure 8(f)) or non-uniformly sampled (see Figure 8(d)). We assess the robustness to model assumptions by simulating signals from the model with ξ_1 following a Gamma or Student t distribution (see Figure 8(g-h)). We also compute the prediction errors under different error distributions (see Figure 8(i)).

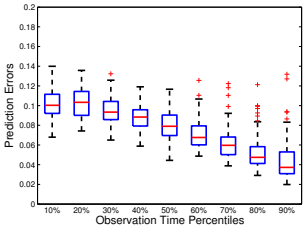
The first observation is that there is insignificant difference in the prediction errors between the true parametric model and the non-parametric degradation model. The differences are larger for high degradation percentiles. Since the difference in the prediction errors increases with additional data we observe for a new component, we infer that this



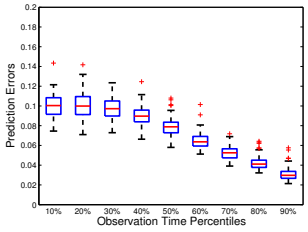
(a) Examples of Degradation Signals



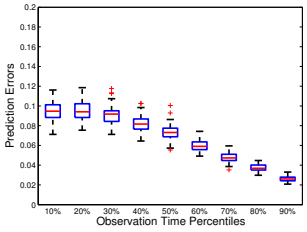
(b) Complete: Non-parametric Model



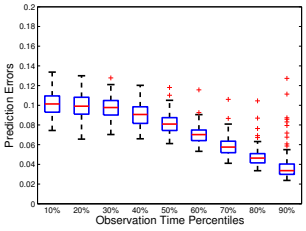
(c) Fragmented: Non-parametric Model using Non-Uniform Sampling



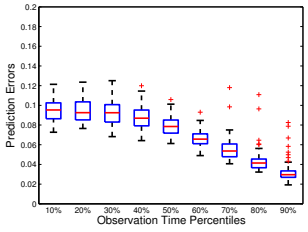
(d) Sparse: Non-parametric Model using Non-Uniform Sampling



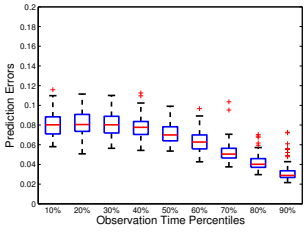
(e) Sparse: True Parametric Model using Non-Uniform Sampling



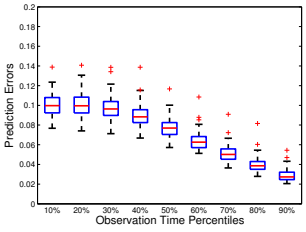
(f) Sparse: Non-parametric Model using Uniform Sampling



(g) Sparse: Non-parametric Model: Gamma Distribution for the Scores



(h) Sparse: Non-parametric Model: t-Distribution for the Scores



(i) Sparse: Non-parametric Model: t-Distribution for the Errors

Figure 8: The prediction error of the residual life estimate for Model 1.

small inefficiency arises due to a decreased accuracy in the estimation of the empirical prior distribution at the later time points.

The second important observation is that the non-uniform sampling technique proposed in Section 2.4 enhances the prediction accuracy of the residual life. In Table 1, we list the median prediction errors based on non-uniform sampling and uniform sampling techniques. The first row of this table represents the time percentile of the degradation signals used for predicting the residual life. It is apparent that the non-uniform sampling technique provides smaller prediction errors, especially at high time percentiles. This is because non-uniform sampling ensures dense coverage of observations over the whole time domain, including the region near maximum observation time(M), and hence provides more accurate estimate of

Table 1: Prediction Errors based on Sparse Degradation Signals that are Uniformly or Non-Uniformly Sampled

Time Percentiles	20%	30%	40%	50%	60%	70%	80%	90%
Uniform	10.08	9.75	9.01	8.17	6.91	5.77	4.79	3.95
Non-Uniform	10.08	9.75	8.97	7.89	6.50	5.28	4.23	3.11

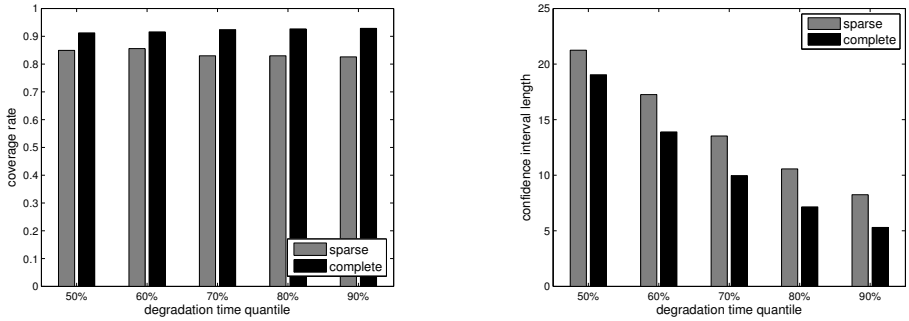
the mean and covariance functions of the model, especially at higher time percentiles.

Lastly, we assess the robustness to departures from our model assumptions: normality of the scores and normality of the errors. In Figure 8 (g-h), we compare the prediction errors when the scores follow Gamma and Student t distribution. We also present the results when the errors follow Student t distribution in Figure 8 (i). The prediction errors for all these different settings are similar. This robustness property of our degradation modeling is inherited from the robustness of the FPCA method [117].

We also evaluate the accuracy of the confidence interval estimates introduced in Section 2.3. In Figure 9, we present the coverage rate level and the mean of the confidence interval length at the degradation lifetime percentiles 50%, 60%, 70%, 80% and 90%. The confidence interval level is $1-\alpha = 0.9$. The coverage rate is higher for complete signals than for sparse signals throughout all percentiles but the difference is insignificant. The coverage rate for both complete and sparse signals is approximately equal to the confidence level $1-\alpha = 0.9$. Moreover, the mean length decreases for higher percentiles implying that the accuracy of the residual life estimate increases as the latest observation time point t^* is closer to the failure time.

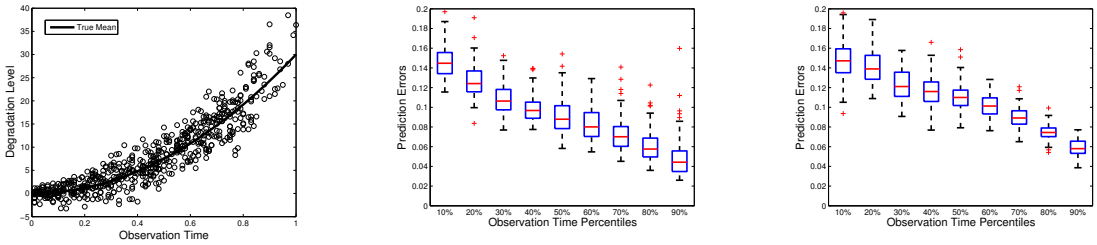
2.6.1.2 Results and Analysis of Model 2

In the following analysis, we still use model 1 as the assumed parametric model and its derived residual life distribution as the benchmark. This assumed parametric model correctly captures the mean degradation trend of model 2 but not the underlying covariance structure of the degradation process. It is worth mentioning that most existing parametric approaches focus on identification of the functional form for the underlying degradation trend ignoring the underlying covariance structure.



(a) Coverage rate (b) Confidence interval length

Figure 9: Confidence interval estimation: the coverage rate (a) and mean length (b). In each plot, the left and the right bars correspond to the sparse and complete scenarios, respectively



(a) Examples of Degradation Signals (b) Sparse: Non-parametric Model (c) Sparse: Parametric Model

Figure 10: The prediction error of the residual life estimate for Model 2.

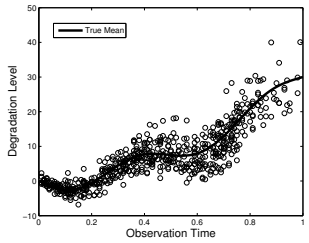
The results in Figure 10 indicate that our non-parametric model is more accurate than the assumed parametric model in predicting the residual life. This is because our proposed non-parametric approach, which is FPCA-based, can not only estimate the mean trend accurately but also capture the dominant modes of the covariance structure correctly. In contrast, parametric models are not flexible enough to accurately capture the underlying covariance structure.

We also compute the prediction error results for cases when the observed degradation signals are complete, fragmented or sparse, and also when the scores and errors follow different distributions. Detailed results can be found in the Appendix.

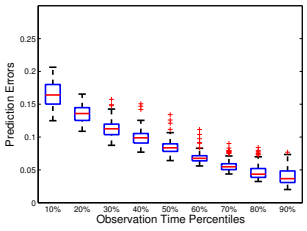
2.6.1.3 Results and Analysis of Model 3

The results are illustrated in Figure 11. In Figure 11(a), we show the simulated degradation observations sparsely sampled from different signals. The thick line in this plot represents the true mean degradation trend, which is non-monotone. In Figure 11(b-d), we compare

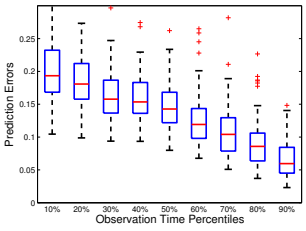
the prediction errors for the non-parametric model under complete, sparse and fragmented scenarios. The low prediction errors indicate the flexibility of our model to apply to situations with non-monotonic degradation signals. Figure 11(e) shows the prediction errors when assuming that Model 1 is the underlying true parametric model of the degradation process (the results are based on sparse degradation signals). Figure 11(f-h) present the results of our model when its model assumptions are violated.



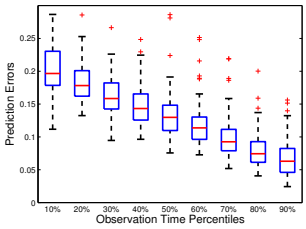
(a) Examples of Degradation Signals



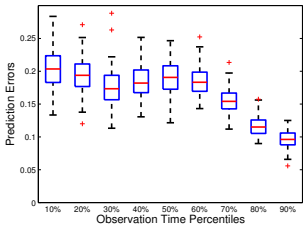
(b) Complete: Non-parametric Model



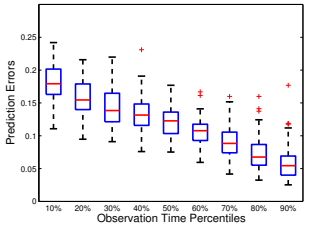
(c) Fragmented: Non-parametric Model



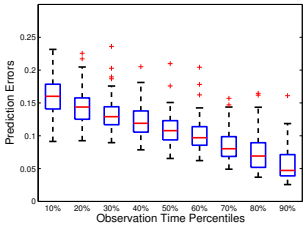
(d) Sparse: Non-parametric Model



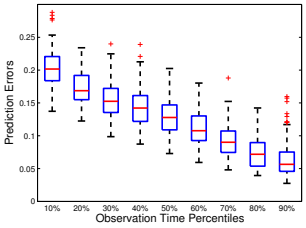
(e) Sparse: Parametric Model



(f) Sparse: Non-parametric Model: Gamma Distribution for the Scores



(g) Sparse: Non-parametric Model: t-Distribution for the Scores



(h) Sparse: Non-parametric Model: t-Distribution for the Errors

Figure 11: The prediction error of the residual life estimate for Model 3.

CHAPTER III

MODELING TRUNCATED DEGRADATION SIGNALS BASED ON FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS

This chapter presents a novel non-parametric degradation modeling framework for predicting and updating, in real-time, the RLDs of partially degraded systems using in-situ truncated degradation signals. Our framework is based on FPCA. Our modeling framework is well-suited for modeling sparse and fragmented degradation signals in addition to complete signals. From a modeling standpoint, one of the primary contributions of this approach is the signal transformation procedure which allows us to apply FPCA to characterize degradation signals that are truncated at the failure threshold. The benefits of using our proposed model are demonstrated using a simulation study as well as crack growth experimental data. We show that our proposed non-parametric model outperforms some of the existing popular parametric and non-parametric models in terms of accurately predicting the remaining lifetime.

We organize this chapter as follows: in Section 3.1, we introduce a non-parametric degradation modeling framework using FPCA. A signal transformation procedure is developed in Section 3.2 to enable us to use FPCA for modeling truncated degradation signals. In Section 3.3, the performance of our modeling approach is evaluated using simulated degradation signals and a real-world crack growth data set.

3.1 Degradation Modeling Using FPCA

3.1.1 Model Development

We denote the amplitude of the degradation signal of unit i as $S_i(t)$ for $i = 1, \ldots, n$, where n represents the number of units, and t represents the observation time in a bounded time domain $[0; M]$, where M is the maximum observation time point. The value of M may represent, for example, the maximum failure time of a system. Since any industrial

application has a finite time-to-failure, M will also be considered to be finite. Next, we decompose the degradation signal, $S_i(t)$, as follows:

$$S_i(t) = \tilde{\mu}(t) + \tilde{X}_i(t) + \tilde{\epsilon}_i(t) \tag{9}$$

where $\tilde{\mu}(t)$ is the underlying trend of the degradation process and is assumed to be deterministic; $\tilde{X}_i(t)$ represents the deviation from the underlying degradation trend and is assumed to be stochastic with mean zero and covariance $\tilde{C}(t,t')$; and $\tilde{\epsilon}_i(t)$ are independent and identically distributed error terms with mean zero and variance $\tilde{\sigma}^2$. $\tilde{X}_i(t)$ and $\tilde{\epsilon}_i(t)$ are assumed to be independent.

Using the Karhunen-Loéve decomposition [47, 66], $\tilde{C}(t,t')$ can be expanded as follows:

$$\tilde{C}(t,t') = \sum_{k=1}^{\infty} \tilde{\lambda}_k \tilde{\phi}_k(t) \tilde{\phi}_k(t'),\ t,t' \in [0;M]$$

where $\tilde{\phi}_k(t)$ for $k = 1, 2, \ldots$ are the associated eigenfunctions with support $[0;M]$ and $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots$ are the ordered nonnegative eigenvalues. Since the eigenfunctions are orthogonal and form a functional basis in the domain $[0;M]$, $\tilde{X}_i(t)$ can be expressed as follows:

$$\tilde{X}_i(t) = \sum_{k=1}^{\infty} \tilde{\xi}_{ik} \tilde{\phi}_k(t) \tag{10}$$

where $\tilde{\xi}_{ik}$ for $k = 1, 2, \ldots$ are called *scores*, and they are independent random variables with means $\mathbb{E}(\tilde{\xi}_{ik}) = 0$ and variances $\mathbb{E}(\tilde{\xi}_{ik}^2) = \tilde{\lambda}_k$. Generally, only a small number of eigenvalues are significantly non-zero. Consequently, we truncate equation (10) at some $\tilde{K}$, where $\tilde{K}$ is the number of significantly non-zero eigenvalues. In this chapter, $\tilde{K}$ is selected such that it minimizes the cross-validation score defined in Equation (10) of [117]. One can also adapt AIC type criteria here, which is much more computationally efficient and the results are still similar to those obtained by cross-validation. The final model is expressed as follows:

$$S_i(t) = \tilde{\mu}(t) + \sum_{k=1}^{\tilde{K}} \tilde{\xi}_{ik} \tilde{\phi}_k(t) + \tilde{\epsilon}_i(t) \tag{11}$$

Within the above construct, the observed degradation signal, $S_i(t)$, follows a stochastic process with mean $E(S_i(t)) = \tilde{\mu}(t)$ and covariance $Cov(S_i(t), S_i(t')) = \tilde{C}(t,t') + \tilde{\sigma}^2 I$. The

mean function and the covariance surface can be estimated from collective data of degradation signals following common non-parametric regression methods, such as splines in [42] or local polynomial smoothers in [117].

3.1.2 Model Estimation with Truncated Degradation Signals

The above FPCA-based framework provides a characterization of degradation signals under the assumption that the mean and covariance functions are estimated accurately using historical degradation signals. However, for truncated degradation data, the estimates of the mean and covariance functions are based on a sample of degradation observations that are selected with bias, i.e., only observations below the failure threshold are observed and used for estimation. The biased selection of the degradation observations in turn results in biased estimates of the mean and covariance functions. Inaccurate estimation of the mean and covariance of the degradation signal leads to inaccurate predictions of the residual lifetime and any measure of the degradation process.

For illustrative purposes, we provide one example in Figure 12. In the left plot, the dots represent the actual observations. The black solid and black dash lines represent the true underlying degradation curve and the estimated mean degradation trend, respectively. We observe that the mean degradation trend is underestimated at later time points. This is due to the biased selection of data, i.e., the degradation signals above the failure threshold are truncated. In the right plot, the solid line represents the actual degradation signal for one unit and the dashed line represents its estimated degradation trend using the FPCA framework. This example demonstrates that the FPCA-based framework described above cannot be used to accurately model truncated degradation signals.

To overcome this challenge, we propose a signal transformation approach that allows us to remedy the shortcomings of applying the FPCA framework.

3.2 Signal Transformation

In this section, we propose a signal transformation procedure that allows us to better utilize FPCA to model truncated degradation signals and predict the RLDs of partially degraded systems that are still operating in the field.

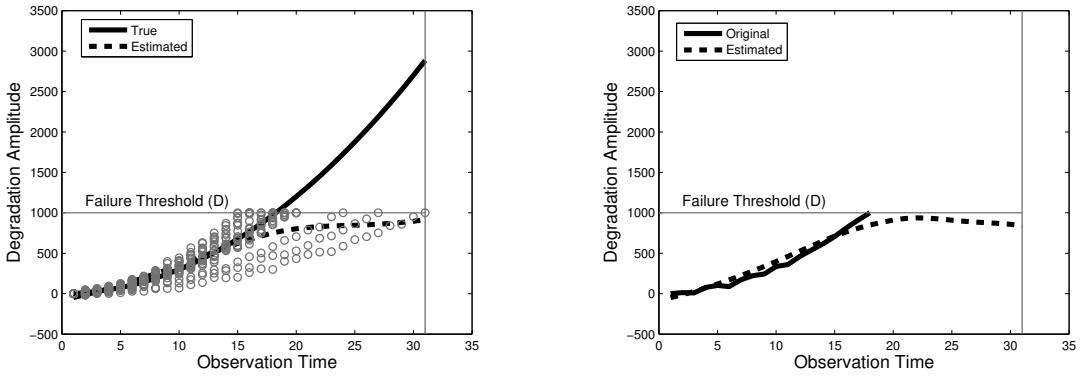


Figure 12: Model Estimation with Truncated Degradation Signals

Traditional degradation models commonly characterize the evolution of the amplitude/level of a degradation signal over time. Typically, the x-axis represents observation time, and the y-axis corresponds to amplitude of a degradation signal up to a maximum D, which is a hard failure threshold (for applications with truncated degradation signals). In contrast, we develop our model to characterize the evolution of the observation time over degradation amplitudes. In other words, we transform degradation signals by axis rotation. After transformation, the x-axis now represents degradation amplitude and the y-axis represents observation time.

We provide an example to illustrate the transformation procedure. In Figure 13, two degradation signals are transformed from the original time domain (left plot of Figure 13) to the new amplitude domain (right plot of Figure 13). The solid lines represent the underlying degradation curves and the dots represent the actual noisy observations.

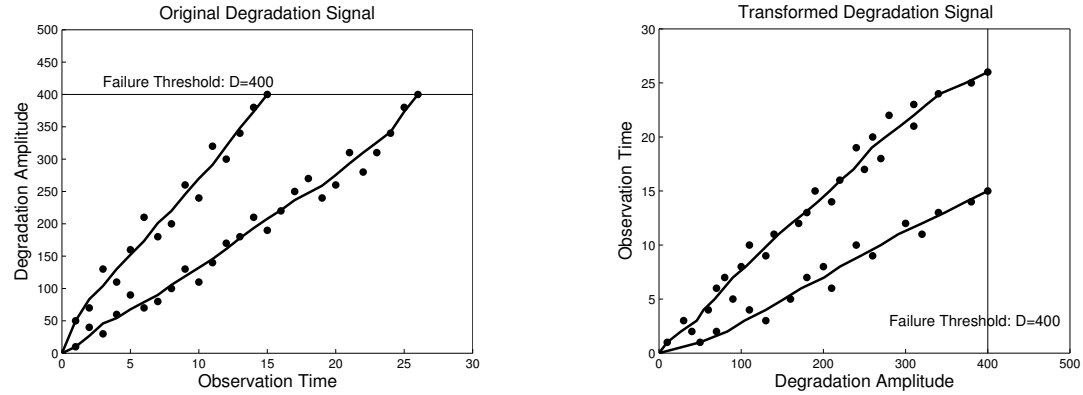


Figure 13: Signal Transformation

Remarks:

1. We assume that the underlying degradation trend is monotonic over time, which is common for degradation data as degradation is a cumulative process. This assumption ensures that the inverse function of the degradation trend exists.

2. After transformation, degradation signals that are truncated in the original time domain are no longer truncated with respect to the new amplitude domain. Consequently, we can estimate the mean and covariance functions accurately based on the transformed degradation signals. This allows FPCA to provide an accurate characterization of the degradation trend in the new domain. It's important to mention that after transformation, the mean and covariance functions to be estimated, the degradation trend to be recovered are all with respect to the new domain. In other words, the transformation technique is not proposed to estimate the mean and covariance functions or recover the degradation trend for the original domain.

3. Many existing methods resort to various assumptions about the distribution of the stochastic parameters in the path model (see [5, 36, 68]) and/or numerical computations (see [67, 84]) to derive approximations of the RLD. In contrast, one advantage of using the axis transformation in Section 3.2.1 is that it allows the derivation of an exact expression of the residual life distribution, which will be shown later in subsection 3.2. After transformation, the problem of determining the lifetime distribution reduces to predicting the value of the underlying degradation process (in the transformed domain) at the failure threshold D.

3.2.1 Modeling of Transformed Degradation Signals

We define $T(d)$ as a component's transformed degradation signal at the degradation amplitude d, where $d \in [0; D]$ and D is the failure threshold. Next, we decompose the transformed degradation signal of component i as follows:

$$T_i(d) = \mu(d) + X_i(d) + \epsilon_i(d), d \in [0; D] \tag{12}$$

where $\mu(d)$ is the underlying trend of the population assumed to be deterministic, $X_i(d)$ represents the deviation from the underlying trend, and $\epsilon_i(d)$ are independent and normally distributed error terms with mean zero and variance σ^2. We also assume $X_i(d)$ and $\epsilon_i(d)$ are independent. Note that equation (12) is similar to but different from equation (9) because equation (9) models the amplitude of the degradation signal over time t, whereas equation (12) models the observation time across varying degradation amplitudes d. By decomposing $X_i(d)$, we can represent equation (12) as follows:

$$T_i(d) = \mu(d) + \sum_{k=1}^{K} \xi_{ik}\phi_k(d) + \epsilon_i(d), d \in [0, D] \tag{13}$$

where the scores ξ_{ik} for $k = 1, 2, \ldots$ are assumed to be independent random variables with mean $\mathbb{E}(\xi_{ik}) = 0$ and variance $\mathbb{E}(\xi_{ik}^2) = \lambda_k$.

Within the above construct, the mean function and the covariance surface can be estimated from the collective data of transformed degradation signals by following common non-parametric regression methods, such as local polynomial smoothers with a constant bandwidth as in [117] or a nearest neighborhood bandwidth as in the R package "locfit" (see [65]).

Under the transformation framework, we estimate the mean and the covariance of $T(d)$ where d is the degradation amplitude. Since d commonly does not follow a Uniform distribution over $[0; D]$, there may be neighborhoods of d where no observation of the process $T(d)$ exists. Using an adaptive bandwidth selection method (e.g. the nearest neighborhood bandwidth in [65]), we select varying bandwidths depending on the density of the observed degradation amplitude d within each neighborhood of d.

In many degradation applications, the mean degradation function $\mu(.)$ is known to be monotonic. This property may not be preserved using the local polynomial method. To ensure monotonicity, one can apply the standard Pool-Adjacent-Violator-Algorithm (PAVA) to the estimated mean function that is obtained using the local polynomial method. PAVA is a simple but efficient iterative algorithm, which provides the best estimator of the original function in terms of the least square standard under monotonicity constraints. More details about PAVA can be found in [7] and [19].

Once the model has been estimated, it can be used to predict the remaining life distribution (RLD) of components that are still operating in the field. In the following section, we will present a Bayesian approach that allows us to continuously update the RLD of an individual fielded component by using its *in-situ* degradation observations.

3.2.2 Bayesian Updating of Non-parametric Degradation Model

Assume that at time t_J, we have observed a partial degradation signal, $S^*(t_1), \ldots, S^*(t_J)$, from a fielded component. The corresponding transformed signal will also be composed of J observations, $T^*(d_1), \ldots, T^*(d_J)$. We are interested in using the observed signal to update the component's RLD. To achieve this, we propose an empirical Bayes method to update the non-parametric degradation model (given by equation (13)) with the component's degradation signal.

Details of the updating process are given in Proposition 1 and Proposition 2. In Proposition 1, we use the corresponding transformed degradation signal to update the scores $(\xi_1^*, \ldots, \xi_K^*)$ that were initially estimated from historical degradation signals. We assume the prior distributions of the scores are normal. Following Proposition 1, Proposition 2 gives a closed-form expression of the updated RLD for this new fielded component. A point estimate can be obtained by taking the median of the RLD derived. The proof for Proposition 1 and Proposition 2 is available upon request.

Proposition 1. *Given the observed degradation signals, $S^*(t_1), \ldots, S^*(t_J)$, and assume that the fielded component's transformed degradation signal $T^*(d)$ follows;*

$$T^*(d) = \mu(d) + \sum_{k=1}^{K} \xi_k^* \phi_k(d) + \epsilon_d$$

and the prior distribution of ξ_k^* is $N(0, \lambda_k)$ with $\xi_1^*, \ldots, \xi_K^*$ independent, the posterior distribution of $(\xi_1^*, \ldots, \xi_K^*)'$ is

$$(\xi_1^*, \ldots, \xi_K^*)' \sim N\left(VC, V\right) \textit{ where}$$

$$
\begin{aligned}
C &= \sum_{j=1}^{J} \frac{T^*(d_j) - \mu(d_j)}{\sigma^2} \Phi_j \ \textit{and} \\
V &= (\frac{1}{\sigma^2} \sum_{j=1}^{J} \Phi_j (\Phi_j)' + B^{-1})^{-1} \ \textit{with} \\
\Phi_j &= (\phi_1(d_j), \ldots, \phi_K(d_j))' \ \textit{and} \\
B &= diag(\lambda_1, \ldots, \lambda_K)
\end{aligned}
$$

Proposition 2. *Given the posterior distribution of the scores $\xi_1^*, \ldots, \xi_K^*$ in Proposition 1, the residual life distribution updated using the degradation signal observed up to time t_J is,*

$$
P(t \leq y | S^*(t_1), ...S^*(t_J), t \geq 0) = \frac{\Phi\left(g^*(y)\right) - \Phi\left(g^*(0)\right)}{1 - \Phi\left(g^*(0)\right)}
$$

where Φ is standard normal cdf and :

$$
\begin{aligned}
g^*(y) &= \frac{y + t_J - \mu^*(D)}{\sqrt{V^*(D)}} \ \textit{with} \\
\mu^*(D) &= \mu(D) + (VC)' \Phi^D \\
\Phi^D &= (\phi_1(D), \ldots, \phi_K(D))' \ \textit{and} \\
V^*(D) &= \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} V_{k1,k2} \phi_{k_1}(D) \phi_{k_2}(D) + \sigma^2
\end{aligned}
$$

3.3 Case Studies

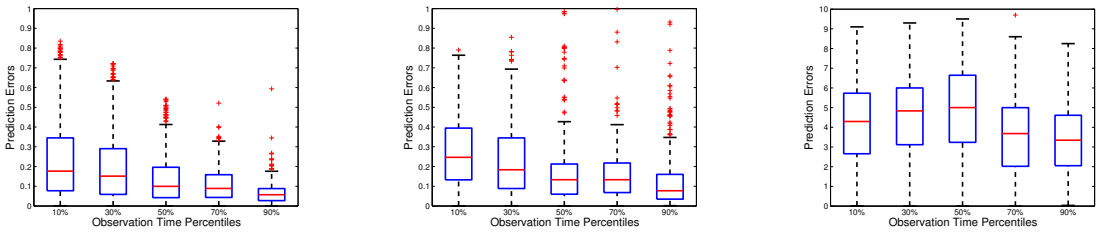
In this section, we evaluate the performance of our degradation framework through two case studies. In the first case study, we focus on comparing our methodology with two alternative non-parametric approaches using simulated truncated degradation signals. We consider only complete signals in this case study. In the second case study, we compare our model with two parametric models based on a real-world crack growth data set. The model performance is investigated more thoroughly, for both complete and incomplete scenarios.

3.3.1 Case Study I: Simulated Degradation Signals

The simulation study presented in this section is designed to evaluate the accuracy of our proposed model with the transformation technique in predicting the residual life with simulated degradation signals. The simulation process proceeds as follows: 200 degradation signals are simulated from the model given by equation (9) with $\mu(t) = 100t^2e^{t^2/9}$, $X_i(t) = \xi_1\phi_1(t)$, where $\xi_1 \sim N(0, 20^2)$, $\phi_1(t) = t^2$, $0 \leq t \leq 1$, and $\sigma = 0.2$. We randomly choose 100 of the simulated signals as the training signals and another 100 as the validation signals. The training signals represent a database of historical degradation signals for estimating the model. On the other hand, the validation signals represent signals that are observed from fielded components. The simulated degradation signals are truncated, and the failure threshold D is assumed to be 2. For a complete signal, we simulate observations every 0.02 time unit. The maximum observation time M is assumed to be 1 time unit. To assess the prediction accuracy, we use the following error criteria:

$$\text{Prediction Error} = \frac{|\text{Predicted Life - Actual Life}|}{\text{Actual Life}} \tag{14}$$

For each validation signal, the prediction errors are reported at varying percentiles of its entire life: 10%, 30%, ..., 90% (90% implies that 90% of the unit's life has passed). The entire simulation process described above is replicated 50 times.



(a) Complete Signal: Our approach

(b) Complete Signals: the "Muller" approach

(c) Complete Signals: the Conventional FPCA approach

Figure 14: Prediction Errors of residual life using three different non-parametric approaches.

We compare our proposed non-parametric approach with two benchmarks. The first

benchmark is the FPCA approach, i.e., without using the transformation technique, hereafter, referred to as "Conventional FPCA". The second benchmark is a model developed by [71], hereafter, referred to as "Müller". As mentioned in the introductory section, the "Müller" approach assumes the lifetime is known for all training units. Therefore, the "Müller" approach is not suitable for applications with sparse or fragmented degradation signals. Consequently, we focus on analyzing the results based on complete signals in this simulation study. The results of using our proposed approach, the "Müller" approach and the "Conventional FPCA" approach are reported in Figure 14(a-c), respectively.

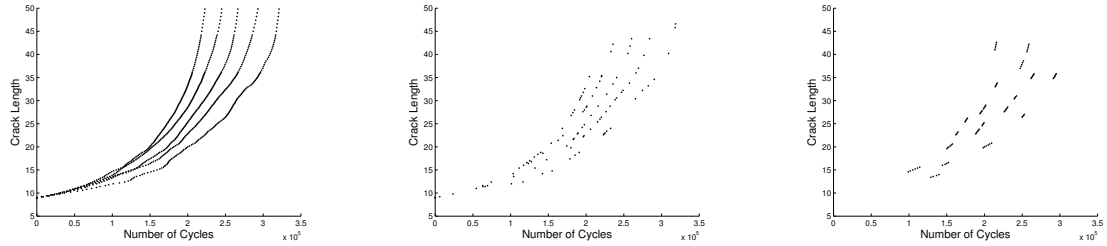
Studying Figure 14, our first observation is that the median and variance of prediction errors in our approach are both smaller than their counterparts in the two benchmarks. For example, at the 90^{th}% percentile of lifetime, the median prediction error of our approach is around 5% and that of the "Müller" approach and the "Conventional FPCA" approach are around 10% and 300%, respectively. One of the disadvantages of the "Müller" approach is that the mean and covariance functions are estimated using reduced training data. More specifically, only the observed trajectories of the units that have survived up to the current time are used in the estimation. The "Müller" approach is also computationally intensive because the whole model estimation process needs to be repeated for every time point. The "Conventional FPCA" approach provides the worst performance due to inaccurate estimates of the mean and covariance functions with the degradation signals being truncated at the failure threshold.

From Figure 14(a), we can observe that the median and variance of the prediction errors decrease with increasing life percentiles (e.g., the median prediction error corresponding to the 10th% percentile is around 20% while the median prediction error corresponding to 90th% percentile is only around 5%), which indicates an improvement of prediction accuracy and precision. The improvement is due to our proposed Bayesian method, which allows updating the score distributions using real-time data from a fielded component. This results in a better estimate of the components' residual life.

3.3.2 Case Study II: Crack Growth Data

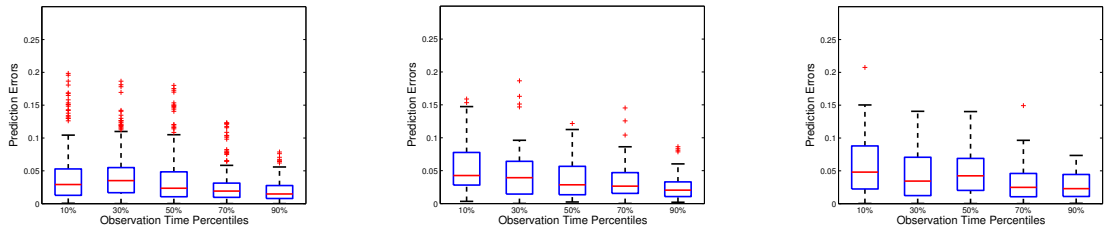
In this section, we apply our proposed non-parametric degradation modeling framework to a crack growth data set (Virkler, 1979). This data set is generated from 68 identical center-cracked panels. Replicate tests are performed under identical conditions. Data is recorded at consistent discrete crack length levels, from 9.0mm to a final length of 49.8mm, which is the failure threshold. Each signal consists of 164 data points.

Of the 68 crack growth degradation signals, we randomly choose 58 as training signals to estimate the degradation model. The rest are validation signals for evaluating the prediction accuracy of our methodology. Each unit is tested until failure (when the crack length reaches a length of 49.8 mm). Therefore, the actual lifetime of each unit is observed. The percentage error given by equation (26) is used to assess the prediction accuracy. For each crack, the prediction errors are reported at the following percentiles of a crack's lifetime: 10%, 30%, ..., 90%. To evaluate how well our approach performs for this data set, this process was replicated 100 times.



(a) Complete Degradation Signals (b) Sparse Degradation Signals (c) Fragmented Degradation Signals

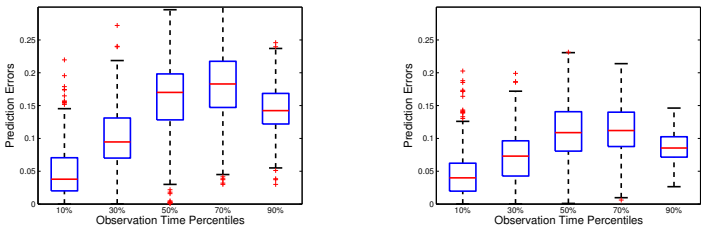
Figure 15: Examples of degradation signals.



(a) Complete: Non-parametric Model (b) Sparse: Non-parametric Model (c) Fragmented: Non-parametric Model

Figure 16: Prediction Errors of residual life using our Non-parametric model

First, we compare the performance of our model under different signal scenarios, i.e.,



(a) Sparse: Log-linear Model (b) Sparse: Log-log-linear Model

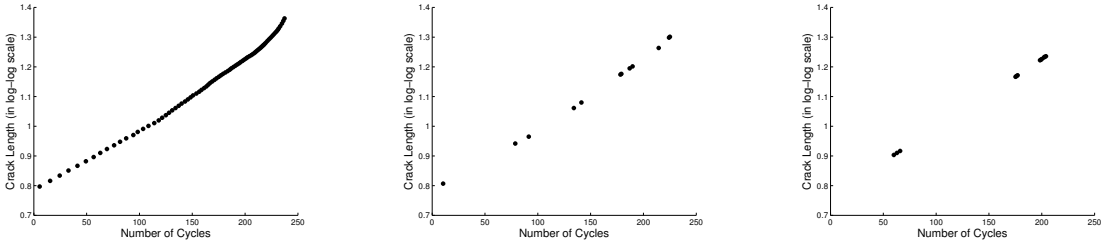
Figure 17: Prediction Errors of residual life using Parametric models

complete, sparse, and fragmented. For the sparse scenario, we uniformly sample 12 observations from each complete degradation signal (in both training and validation data sets). For the fragmented case, we uniformly sample 3 fragments from each complete degradation signal (in both training and validation data sets) with lengths uniformly generated from $\{3, 4, 5\}$. A representative sample of complete, sparse and fragmented degradation signals is shown in Figure 15. The results of applying our model to complete, sparse and fragmented degradation signals are illustrated in Figure 16 (a-c), respectively. In each plot, we observe that the median prediction errors corresponding to each percentiles are all less than 5%. This indicates that our nonparametric model performs very well for complete as well as incomplete degradation signals. Besides, the performance is better at higher lifetime percentiles, which once again demonstrates the benefits of our Bayesian updating framework. Another observation is that the prediction errors are smaller for sparse signals than for fragmented signals. A possible explanation is that the sparse data is more spread out giving us a better chance to get information of units at different stages of their life span. Therefore, the resulting estimates or predictions might be slightly more accurate.

To benchmark our methodology, we compare our proposed non-parametric degradation model with two parametric models. Since the degradation signals exhibit an exponentially increasing trend, we transform the degradation signals using a natural logarithm in order to linearize the trend and then apply the linear random coefficients model proposed in [36] (hereafter, referred to as "Log-linear"). We also consider a double logarithm transformation of the original degradation data (hereafter, referred to as "Log-log-linear") and apply the same model. The results of utilizing these two parametric models with sparse degradation

signals are reported in Figure 17. It is clear that both parametric models perform relatively worse than our non-parametric approach in terms of predicting the residual life. We believe this is because these types of models are not flexible enough to capture the degradation trend exhibited in these degradation signals. To demonstrate our belief, we provide one example in Figure 18 to illustrate the bias induced by the parametric model "Log-log-linear".

In Figure 18, the x-axis represents degradation time and the y-axis represents degradation amplitude, but in the log-log scale. We have one complete, sparse and fragmented degradation signal in Figure 18(a-c), respectively. If the "Log-log-linear" model is the true underlying parametric model, we should see a linear trend in all three plots. This seems to be true in the sparse and fragmented scenarios (see Figure 18(b-c)). However, for the complete signal in Figure 18(a), we observe that the degradation trend is still nonlinear, for example, the degradation rate at the end of the signal is faster than the other parts. This indicates that the log-log-transformation does not linearize the signal trend (the same applies for the "Log" transformation). Therefore, the "Log-log-linear" model does not accurately capture the crack propagation trend throughout the subject's lifetime. This example shows the potential difficulty of identifying a reasonable parametric model for sparse and fragmented degradation signals and in turn, demonstrates that our non-parametric model is more robust than parametric models in terms of model mis-specification.



(a) An Example of Complete Signal(in log-log-scale)

(b) An Example of Sparse Signal(in log-log-scale)

(c) An Example of Fragmented Signal(in log-log-scale)

Figure 18: Examples of the crack data under the log-log scale.

CHAPTER IV

MODELING TRUNCATED DEGRADATION SIGNALS BASED ON FUNCTIONAL TIME WARPING ANALYSIS

In this chapter, we propose a functional time warping model for characterizing truncated degradation signals. Our framework is built upon the assumption that all systems follow an identical degradation trend, but they may progress at different degradation rates, which are characterized by random time warping processes. To ensure model flexibility, we use a non-parametric decomposition of the time warping processes by representing them with a set of basis functions. This general structure allows us to model a group of signals that may have different shapes. Using this model, we can predict the RLD of a fielded system using partial observations of its truncated degradation signal. The benefits of our proposed model are demonstrated by using bearing degradation data and simulated degradation signals.

The remainder of the chapter is organized as follows. We first describe the general problem and its challenges in Section 4.1. We present the time warping model in Section 4.2 and apply this model to prediction of the residual life of fielded systems in Section 4.3. The performance of our model is investigated using real-world bearing degradation signals in Section 4.4 and simulated degradation signals in Section 4.5. Some technical details and additional simulation studies are deferred to the Appendix B.

4.1 Problem Description and Challenges

We denote a component's degradation process by $X(t)$, where t is the chronological time. In our framework, a component's underlying failure time T_u is defined as the time when its degradation process $X(\cdot)$ reaches a failure threshold D. Depending on the applications, the failure threshold may be above or below the starting point. Without loss of generality, we focus on the former case. Under this assumption, T_u is defined as follows:

$$T_u = \inf_{t \geq 0}[X(t) \geq D]. \tag{15}$$

Ideally, $X(\cdot)$ could be observed continuously over time and with no measurement errors. In practice, a degradation process can only be monitored at discrete design points, and is commonly contaminated with measurement errors, denoted by $\epsilon(\cdot)$. We represent the observed degradation signal by $Y(t) = X(t) + \epsilon(t)$, where t is a vector of discrete time points. Without loss of generality, we assume that the degradation signal is collected at $t = (d, 2d, 3d, \ldots)$, where d represents the observation time interval. We note that degradation signals do not need to be observed regularly.

In applications with truncated signals, degradation processes are monitored up to the observed failure time T_o, which is defined as the time when the observed degradation signal $Y(\cdot)$ first reaches the failure threshold D:

$$T_o = \min_{t=md, m \in N^+}[Y(t) \geq D]. \tag{16}$$

Therefore, T_u is <u>not</u> observed, but instead its approximation T_o is available.

To illustrate the definitions of T_u and T_o, we provide an example in Figure 19. In this figure, the solid line represents the degradation process $X(\cdot)$ and the solid points represent the discrete degradation observations of $Y(t)$. The degradation signal is truncated at the failure threshold $D = 12$. According to the definitions in (15-16), $T_u = 3.6$ and $T_o = 4$.

In this paper, our primary objective is to predict the underlying residual life of a fielded component, denoted by RL_u^* ($RL_u^* = T_u^* - t^*$), using the partial observations $Y^*(t)$ of its degradation signal, observed up to a certain time point t^*. To this purpose, we model T_u as a function of the component's observations and a set of (unknown) parameters. Particularly, we specify a statistical model G, such that $T_u = G(Y(t), \boldsymbol{\beta})$, and then estimate the model parameters $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$ (In this paper, vectors and matrices are displayed as bold). This will provide a prediction for RL_u^* as follows:

$$\hat{RL}_u^* = \hat{T}_u^* - t^* = G(Y^*(t), \hat{\boldsymbol{\beta}}) - t^*$$

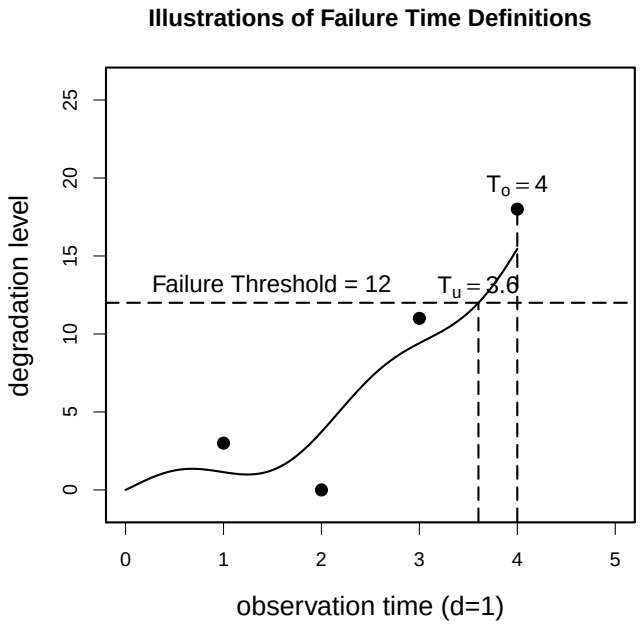


Figure 19: An example of T_u vs. T_o.

Most methods require observing T_u in the training stage. As mentioned above, one can approximate T_u given the observed truncated signal. A direct approximation would be $\tilde{T_u} = T_o$. The bias due to the approximation of T_u depends on both the underlying process and the parameters relevant to the observation process, specifically, the observation time interval d and the amplitude of the measurement error $\epsilon(\cdot)$. It is important to devise an estimation approach for $G(Y(t), \boldsymbol{\beta})$ that is not sensitive to this approximation.

One approach for deriving the statistical model $G(Y(t), \boldsymbol{\beta})$ is using regression analysis in which the underlying failure time is the response variable, and the predictors consist of the observed degradation signals. The regression model parameters may be estimated using training degradation signals. Given $Y^*(t)$, the observed signal of a fielded component, one can then predict its underlying failure time by using the trained regression model. For example, in [71], the authors propose a time-varying functional regression method for predicting the remaining life of medflies. In their approach, they apply a functional principal component analysis to the signals up to the current time, and therefore, estimating a time-varying model. The observed functional signals are then represented by the time-varying functional principal component scores, which are continuously updated as time progresses.

The time-varying scores become the predictor variables in the functional regression model.

In this paper, we first estimate a model for $X(t)$ and use the definition in (15) along with the model for $X(t)$ to obtain predictions for T_u. This first step provides a prior model distribution for $X(t)$. Similarly to our previous research in [36, 120, 119], we update the prior model distribution of $X(t)$ using the partial observations $Y^*(t)$ of a fielded component, denote this updated model distribution $X^*(t)$, and use it further to predict T_u^*. This paper introduces a novel approach for the model distribution of $X(t)$ given that we observe truncated degradation signals. In Section 4.2, we present the model decomposition and assumptions for $X(t)$. One advantage of our modeling approach is that it provides a direct prediction formula for T_u^*. We present this in Section 4.3.

4.2 Modeling and Monitoring Truncated Signals

In this section, we present a time warping framework for modeling the underlying degradation process $X(t)$ based on truncated degradation signals. The associated model parameters are estimated using a Bayesian approach. This is the first step in our approach as discussed in the previous section.

4.2.1 Model Framework and Assumptions

The training data set consists of a collection of degradation signals $Y_i(t_{ij})$, for $i = 1, \ldots, I$ (I is the number of components) and $j = 1, \ldots, m_i$ (m_i is the number of observation time points for signal i).

The central assumption of the underlying degradation process is $X_i(\cdot) = Z(h_i(\cdot))$, where $Z(\cdot)$ is a shape function that characterizes the common degradation trend across components and $h_i(\cdot)$ is a synchronizing function that characterizes the rate at which the ith component degrades. Under this assumption, the model decomposition is

$$Y_i(t) = X_i(t) + \varepsilon_i(t) = Z(h_i(t)) + \varepsilon_i(t), \quad t \in [0, T_{i,u}] \tag{17}$$

$$Z(s) = \boldsymbol{B_m}'(s)\boldsymbol{\beta}, \quad s \in [0, 1] \tag{18}$$

$$\mu_i(s) \equiv h_i^{-1}(s) = \boldsymbol{B_\mu}'(s)\boldsymbol{\phi_i}, \quad s \in [0, 1] \tag{19}$$

More details and assumptions about this model are as follows:

• Throughout this paper, s will represent synchronized time and t will represent physical observation time, which is discrete in practice. $t \in [0, T_{i,u}]$, where $T_{i,u}$ represents the underlying failure time of the ith unit.

• The measurement error term $\varepsilon_i(t)$ is assumed to be independent and normally distributed with zero mean and variance σ_ε^2. It is also assumed to be independent of all the other random terms in the model.

• The shape function $Z(s)$ is assumed to be unknown, and therefore, decomposed using a set of basis functions $\boldsymbol{B_m}(s)$ (we use cubic B-spline basis functions throughout this paper) of the dimension d_β. We assume $Z(s)$ is smooth and model it using penalized regression splines, where the penalty, which accounts for the differences of adjacent B-spline coefficients, is defined to ensure adequate smoothness of the fitted curves [24]. Under the Bayesian framework [57], this corresponds to specifying a first-order random walk shrinkage prior on the shape coefficients $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_{d_\beta})$ such that

$$\beta_k = \beta_{k-1} + e_k, \quad e_k \sim N(0, \lambda_\beta)$$

Under this formulation, $\boldsymbol{\beta} \sim N(0, \boldsymbol{\Sigma_\beta})$, with

$$\boldsymbol{\Sigma_\beta}^{-1} = \boldsymbol{\Omega_\beta}/\lambda_\beta; \boldsymbol{\Omega_\beta} = \begin{pmatrix} 2 & -1 & 0 & & & 0 \\ -1 & 2 & -1 & \ddots & & \\ 0 & -1 & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -1 & 0 \\ & & & -1 & 2 & -1 \\ 0 & & & 0 & -1 & 1 \end{pmatrix}_{d_\beta \times d_\beta}$$

where $\boldsymbol{\Omega_\beta}$ is a precision matrix and λ_β controls the smoothness level of $Z(s)$.

• The time warping function $\mu_i(s)$ is the inverse of the synchronizing function $h_i(t)$. Since we assume $\mu_i(s)$ has an unknown shape, we decompose $\mu_i(s)$ using a set of basis functions $\boldsymbol{B_\mu}(s)$ of the dimension d_ϕ. The functions $\mu_i(s)$ and $h_i(t)$ are subject to the following two assumptions:

$*$ (monotonicity) $\mu_i(s)$ and $h_i(t)$ are both strictly monotone increasing, and therefore invertible;

Remark: The monotonicity constraint ensures a one-to-one correspondence between synchronized time s and physical observation time t. In our definition, $\mu_i(s) \equiv h_i^{-1}(s)$, or equivalently, $h_i(t) \equiv {\mu_i}^{-1}(t)$. To ensure the monotonicity of the time warping functions, we constrain $\boldsymbol{\phi_i} = (\phi_{i,1}, \cdots, \phi_{i,d_\phi})$ such that

$$\phi_{i,1} < \phi_{i,2} < \ldots < \phi_{i,d_\phi}. \tag{20}$$

Under the above condition for $\boldsymbol{\phi_i}$, the derivative of $\mu_i(s)$ is positive according to the basic properties of B-spline. Therefore, $\mu_i(s)$ is a monotone increasing function. Note that this condition is sufficient but not necessary. Details can be found in [13].

$*$ (image) $0 \leq h_i(t) \leq 1$, $h_i(0) = 0$, $h_i(T_{i,u}) = 1$.

Remark: The standardized time interval $[0, 1]$ has an intuitive interpretation. Every component starts to degrade from the beginning of its lifetime, represented by the synchronized time $s = 0$, to the end of it, represented by the synchronized time $s = 1$. Each component corresponds to a specific time warping function since each component degrades at its own rate. To ensure that the image assumption holds, we constrain $\boldsymbol{\phi_i}$ such that

$$\phi_{i,1} = 0, \quad \phi_{i,d_\phi} = T_{i,u}. \tag{21}$$

• We assume that the time warping functions $\mu_i(s)$ are smooth and model these functions using penalized regression splines by placing a first-order random walk shrinkage prior on the coefficients. Specifically, we assume $\boldsymbol{\phi_i} \sim N(b_0 \boldsymbol{\Upsilon}, \boldsymbol{\Sigma_\phi})$ where $\boldsymbol{\Upsilon}$ is the vector associated with the identity time-transformation function, i.e., $\boldsymbol{B_\mu}' \boldsymbol{\Upsilon} = t$. ${\boldsymbol{\Sigma_\phi}}^{-1} = \boldsymbol{\Omega_\phi}/\lambda_\phi$, where $\boldsymbol{\Omega_\phi}$ is a precision matrix of the dimension d_ϕ that has the same structure as that of $\boldsymbol{\Omega_\beta}$ and λ_ϕ is a scalar that controls the smoothness level.

In the simpler case when $\mu_i(s) = b_0 s$, the time warping process involves only stretching or compressing the time domain at a constant rate. In other words, the time warping process retains the original shape of the common shape function. Here b_0 can be interpreted as an average acceleration factor, which describes the overall mean degradation rate.

We provide one example in Figure 20 to illustrate this modeling framework. In this figure, the long dashed line represents a common shape function, which is quadratic with respect to the synchronized time $s \in [0, 1]$, i.e., $Z(s) = s^2$. The dot-dashed lines represent the underlying degradation processes of two different components. For component 1, the time warping function is linear: $\mu_1(s) = 5s$, therefore, $h_1(t) = \frac{t}{5}$. According to our model, this leads to $X_1(t) = Z(h_1(t)) = (\frac{t}{5})^2$ and the corresponding underlying failure time $T_{1,u} = 5$. For component 2, the time warping function is quadratic: $\mu_2(s) = 3s^2$, therefore, $h_2(t) = \sqrt{\frac{t}{3}}$. According to our model, this leads to $X_2(t) = Z(h_2(t)) = \frac{t}{3}$ and the corresponding underlying failure time $T_{2,u} = 3$.

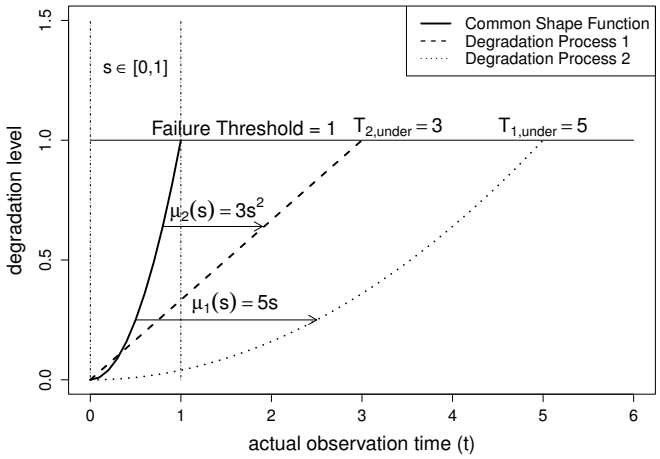


Figure 20: Illustration of two degradation signals following our central time-warping model assumption.

4.2.2 Model Estimation

To simplify the derivation of our model estimation procedure, we divide the model parameters into $\boldsymbol{\theta} = (\sigma_\varepsilon^2, \boldsymbol{\beta}, b_0, \lambda_\phi)$ and $\boldsymbol{\phi} = (\boldsymbol{\phi_1}, \ldots, \boldsymbol{\phi_I})$. One approach to estimating these parameters is by maximizing the marginal log-likelihood of the pooled degradation data

$\boldsymbol{Y} = (\boldsymbol{Y_1}, \ldots, \boldsymbol{Y_I})$ decomposed as:

$$l_Y = \sum_{i=1}^{I} l_{Y_i} = \sum_{i=1}^{I} log \int f(\boldsymbol{Y_i}|\boldsymbol{\phi_i}; \boldsymbol{\theta}) f(\boldsymbol{\phi_i}|\boldsymbol{\theta}) d\boldsymbol{\phi_i} = \sum_{i=1}^{I} log \int f(\boldsymbol{Y_i}|\boldsymbol{\phi_i}; \boldsymbol{\beta}, \sigma_\varepsilon^2) f(\boldsymbol{\phi_i}|\lambda_\phi, b_0) d\boldsymbol{\phi_i} \ .$$

Due to the model complexity, maximizing the above marginal log-likelihood is intractable. In this paper, we resort to a Bayesian approach for model estimation. For computational convenience, we choose conjugate priors for all parameters that are needed for estimation (In our notation, if $X \sim Gamma(k, r)$, then $E(X) = k/r$.):

$$1/\sigma_\varepsilon^2 \sim Gamma(k_{\sigma_\varepsilon}, \gamma_{\sigma_\varepsilon}) \text{ and } 1/\sigma_{b_0}^2 \sim Gamma(k_{\sigma_{b_0}}, r_{\sigma_{b_0}})$$
$$b_0 \sim Normal(m_{b0}, \sigma_{b_0}^2)$$
$$1/\lambda_\beta \sim Gamma(k_{\lambda_\beta}, r_{\lambda_\beta}) \text{ and } 1/\lambda_\phi \sim Gamma(k_{\lambda_\phi}, r_{\lambda_\phi})$$

Given the observed data Y, the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\phi})$ is:

$$f(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{Y}) \propto f(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{\phi}) f(\boldsymbol{\theta}, \boldsymbol{\phi})$$
$$= f(\boldsymbol{Y}|\boldsymbol{\phi}; \boldsymbol{\beta}, \sigma_\varepsilon^2) f(\boldsymbol{\phi}|b_0; \lambda_\phi) f(\boldsymbol{\beta}|\lambda_\beta)$$
$$\times f(\sigma_\varepsilon^2|k_{\sigma_\varepsilon}, r_{\sigma_\varepsilon}) f(b_0|m_{b_0}, \sigma_{b_0}^2) f(\sigma_{b_0}^2|k_{\sigma_{b_0}}, r_{\sigma_b}) f(\lambda_\beta|k_{\lambda_\beta}, r_{\lambda_\beta}) f(\lambda_\phi|k_{\lambda_\phi}, r_{\lambda_\phi})$$

Since the joint posterior density is analytically intractable, we propose to use an MCMC-based posterior inference procedure. More specifically, we implement a Metropolis-within-Gibbs algorithm. In this algorithm, we use Metropolis-Hastings sampling to sample the time-transformation coefficients $\boldsymbol{\phi_i}$, $i = 1, \ldots, I$ and Gibbs sampling to sample all other parameters from their respective closed form full conditionals.

The details about the full conditionals are given as follows. Let θ_j denote the jth element of $\boldsymbol{\theta}$ and let $\boldsymbol{\theta_{-j}}$ denote the vector containing all components of $\boldsymbol{\theta}$ except θ_j. All components in $\boldsymbol{\theta}$ can be derived with closed-form full conditional distributions, $\theta_j|\boldsymbol{\theta_{-j}}, \boldsymbol{Y}$. In the following results, we use t^g to denote the grid vector of the observation time points, i.e., $t^g = (t_1, ..., t_n)'$. $Vec(\boldsymbol{A})$ stacks an $n \times m$ matrix $\boldsymbol{A}$ in a vectorized form so that $Vec(\boldsymbol{A}) = [a_{11}, ...a_{1m}, ...a_{n1}, ...a_{nm}]'$. The full conditional distributions for the common shape parameters $\boldsymbol{\beta}$, the mean acceleration factor b_0 and the variance parameters σ_ε^2, λ_β, λ_ϕ are listed below.

- The full conditional distribution for $\boldsymbol{\beta}$ is:

$$(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\theta}_{-\boldsymbol{\beta}}) \sim N(m_\beta; \boldsymbol{V_\beta})$$

where: $\boldsymbol{V_\beta}^{-1} = (\boldsymbol{\Omega_\beta}/\lambda_\beta) + 1/\sigma_\varepsilon^2 \boldsymbol{X'X}$ and $m_\beta = \boldsymbol{V_\beta}[1/\sigma_\varepsilon^2 \times \boldsymbol{X'Y}]$ with

$\boldsymbol{X} = (Vec[\boldsymbol{B_m}(\mu(t^g; \boldsymbol{\phi_1}))], ...Vec[\boldsymbol{B_m}(\mu(t^g; \boldsymbol{\phi_N}))])$

- The full conditional distribution for b_0 is:

$$(\mathrm{b}_0|\boldsymbol{Y}, \boldsymbol{\theta}_{-\mathrm{b}_0}) \sim N(m^*_{\mathrm{b}_0}, \sigma^{*2}_{\mathrm{b}_0})$$

where: $(\sigma^{*2}_{\mathrm{b}_0})^{-1} = \boldsymbol{\Upsilon}'(\boldsymbol{\Omega_\phi}/\lambda_\phi)^{-1}\boldsymbol{\Upsilon} + 1/\sigma^2_{\mathrm{b}_0}$ and $m^*_{\mathrm{b}_0} = \sigma^{*2}_{\mathrm{b}_0}[\sum_{i=1}^{N} \boldsymbol{\Upsilon}'(\boldsymbol{\Omega_\phi}/\lambda_\phi)^{-1}\boldsymbol{\phi_i} + m_{b_0}/\sigma^2_{\mathrm{b}_0}]$

- The full conditional distribution for σ^2_ε is:

$$(1/\sigma^2_\varepsilon|\boldsymbol{Y}, \boldsymbol{\theta}_{-\sigma^2_\varepsilon}) \sim Gamma(k^*_{\sigma_\varepsilon}, r^*_{\sigma_\varepsilon})$$

where: $k^*_{\sigma_\varepsilon} = k_{\sigma_\varepsilon} + (\sum_{i=1}^{N} n_i)/2$ and $r^*_{\sigma_\varepsilon} = r_{\sigma_\varepsilon} + 1/2 \sum_{i=1}^{N} [\boldsymbol{Y_i} - \boldsymbol{B_m}(\mu(t^g; \phi_i)\boldsymbol{\beta}]'[\boldsymbol{Y_i} - \boldsymbol{B_m}(\mu(t^g; \boldsymbol{\phi_i})\boldsymbol{\beta}]$

- The full conditional distribution for λ_β is:

$$(1/\lambda_\beta|\boldsymbol{Y}, \boldsymbol{\theta}_{-\lambda_\beta}) \sim Gamma(k^*_{\lambda_\beta}, r^*_{\lambda_\beta})$$

where: $k^*_{\lambda_\beta} = k_{\lambda_\beta} + d_\beta/2$ and $r^*_{\lambda_\beta} = r_{\lambda_\beta} + 1/2\boldsymbol{\beta'\Omega_\beta\beta}$

- The full conditional distribution for λ_ϕ is:

$$(1/\lambda_\phi|\boldsymbol{Y}, \boldsymbol{\theta}_{-\boldsymbol{\phi}}) \sim Gamma(k^*_{\lambda_\phi}, r^*_{\lambda_\phi})$$

where: $k^*_{\lambda_\phi} = k_{\lambda_\phi} + d_\phi N/2$ and $r^*_{\lambda_\phi} = r_{\lambda_\phi} + 1/2 \sum_{i=1}^{N} (\boldsymbol{\phi_i} - b_i\boldsymbol{\Upsilon})'\boldsymbol{\Omega_\phi}(\boldsymbol{\phi_i} - b_i\boldsymbol{\Upsilon})$

Details about using Metropolis-Hastings algorithm for estimating $\boldsymbol{\phi_i}$, $i = 1, \ldots, I$ are provided in the Supplementary Material B. The estimates of population-level parameters $\boldsymbol{\beta}, b_0, \sigma^2_\varepsilon, \lambda_\phi$ are obtained by averaging each posterior samples. We denote the resulting estimates by $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{b}_0, \hat{\sigma}^2_\varepsilon, \hat{\lambda}_\phi)$.

The foregoing implementation requires the specification of the tuning parameters d_β and d_ϕ, which reduces to a model selection problem. One could implement cross-validation to select d_β and d_ϕ to minimize a predefined measure for the prediction errors. However, this approach is computationally prohibitive. In this paper, we propose to use the Bayes factor [48] for determining d_β and d_ϕ, as it is consistent with our Bayesian estimation framework and computationally more efficient. More details about using the Bayes Factor for model selection are presented in the Appendix B.

4.3 Residual Life Prediction

The focus of this paper is on predicting the RLD of a fielded component using partial observations of its degradation signal. In this section, we discuss how to derive this prediction using the time warping model for the underlying degradation process $X(t)$.

Assume that we have collected a partial degradation signal $Y^*(t)$ for a fielded component, where t is a vector of discrete time points within the interval $[0; t^*]$. To predict the RLD for this new component, one could simply use the time warping model for $X(t)$ without taking into account the additional observations $Y^*(t)$. We conjecture that the RLD prediction is more accurate if instead we update the distribution of $X(t)$ given the partial degradation signal $Y^*(t)$, denote it $X^*(t)$, and use the formula in equation (15) to predict T_u^*.

An appealing advantage of our proposed time warping approach is that we have a very simple formula for predicting the residual life of a fielded component due to the image assumption:

$$RL_u^* = T_u^* - t^* = \mu^*(1) - t^* = \boldsymbol{B_\mu}'(1)\boldsymbol{\phi}^* - t^*. \tag{22}$$

Therefore, we do not need to obtain the updated distribution model $X^*(t)$ but only the updated basis coefficient of its time warping function $\boldsymbol{\phi^*}$. We resort to MCMC to estimate the distribution of $\boldsymbol{\phi^*}$. Specifically, we use Metropolis-Hastings sampling to sample $\boldsymbol{\phi^*}$ from its posterior distribution,

$$f(\boldsymbol{\phi^*}|\boldsymbol{Y^*}; \hat{\boldsymbol{\theta}}) \propto f(\boldsymbol{Y^*}|\boldsymbol{\phi^*}; \hat{\boldsymbol{\theta}}) f(\boldsymbol{\phi^*}|\hat{\boldsymbol{\theta}}) = f(\boldsymbol{Y^*}|\boldsymbol{\phi^*}; \hat{\sigma}_\varepsilon^2, \hat{\boldsymbol{\beta}}) f(\boldsymbol{\phi^*}|\hat{b}_0, \hat{\lambda}_\phi).$$

Using this method, we obtain a sample from the posterior distribution of RL_u^*. Further, based on this sample, we can obtain an estimate of the residual lifetime by taking the posterior mode or mean. Credible intervals can be used to provide a range of plausible values for the residual lifetime.

4.4 Case Study of Rotating Machinery Degradation

The use of rotating machinery is widespread in industry. A large percentage of maintenance activities associated with this type of machinery generally involves inspection and replacement of bearings. Failure in bearings is generally due to a fatigue process. The

Table 2: Estimates of Model Parameters

Parameters	$\hat{d}_\beta$	$\hat{d}_\phi$	$\hat{\boldsymbol{\beta}}$	$\hat{b}_0$	$\hat{\lambda}_\phi$	$\hat{\sigma}_\varepsilon$
Estimates	4	4	[0, 0.00749, 0.00969, 0.025]	152.8	365	0.00407

degradation process usually starts with spalling (the creation of minute pits) on the surface of the steel raceways. As the rolling elements pass over the spalls, they cause an increase in the vibration levels of the machine. Bearing-specific frequencies can then be observed using vibration monitoring techniques. These frequencies are a function of the rotational speed, type of bearings, number of rolling elements, and bearing geometry.

In our experiment, 34 bearings were tested until failure. For each bearing, vibration data was collected every 2 minutes, about 100 observations per signal. The experimental set up is described in detail by [36]. We compare our proposed approach with the parametric model in [36] as a benchmark, which is proposed to model the same data set. In this model, an exponential functional form is identified to represent the degradation trend and the error term is assumed to follow a Brownian Motion process (we denote this model by "exp-brownian" hereafter). Examples of bearing degradation signals are provided in Figure 3(a).

To evaluate the performance of our proposed degradation model, we use these vibration-based degradation signals and repeat the following study for 50 times. We randomly select 20 signals as training signals and the other 14 degradation signals are used as testing signals for evaluation of the absolute prediction errors. The training signals represent a database of training degradation signals for estimating the model. On the other hand, the validation signals represent signals that are observed from fielded bearings. Figure 21 shows the estimated common shape function; the degradation trend is neither linear nor exponential as assumed by the parametric model in [36]. Table 2 summarizes the model parameter estimates.

For each test bearing, we predict its RLD by using the partial observations of its degradation signal. To assess the prediction accuracy, we use the following error criterion:

$$\text{Absolute Prediction Error} = \frac{|\text{Predicted Life - Actual Life}|}{\text{Actual Life}} \tag{23}$$

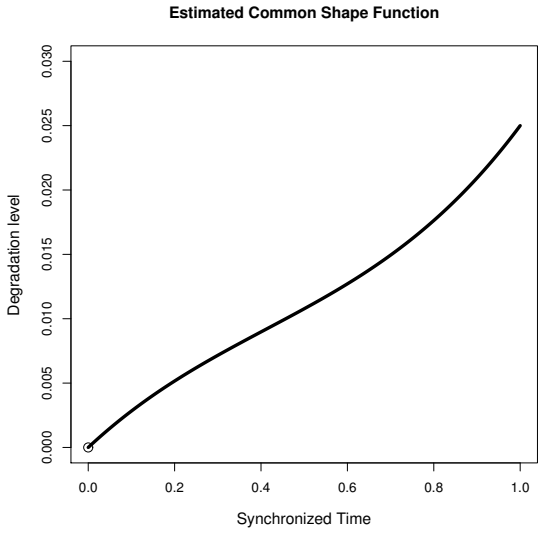


Figure 21: Estimated Common Shape Function

The prediction results of using "time warping" and "exp-brownian" are summarized in Figure 22. The blue solid dots and lines represent the median and (± 2) standard deviations of the absolute prediction errors of "time warping". We first observe that both the median and standard deviations of the absolute prediction errors decrease as the lifetime percentile increases. For example, the median absolute prediction error corresponding to the 10th% percentile is around 30% while the median absolute prediction error corresponding to 90th% percentile is only around 7%. The proposed method thus shows clear improvements as new fielded data is available. We conjecture that the improvements are due to our proposed Bayesian method, which allows updating the distribution of the time warping coefficient by using real-time data.

In Figure 22, the red dashed dots and lines represent the results of "exp-brownian". From this figure, we observe that the median absolute prediction errors of "time warping" are consistently smaller than those of "exp-brownian" at all lifetime percentiles. This indicates that our proposed time warping approach predicts the RLD of fielded bearings with higher accuracy. At all lifetime percentiles except 50%, the standard deviations of the absolute prediction errors are smaller for those of "time warping" than for "exp-brownian". This indicates that our proposed time warping approach is more stable in predicting the RLD of fielded bearings. The superior performance of our proposed time warping approach

demonstrates that our model is relatively more robust to model mis-specifications.

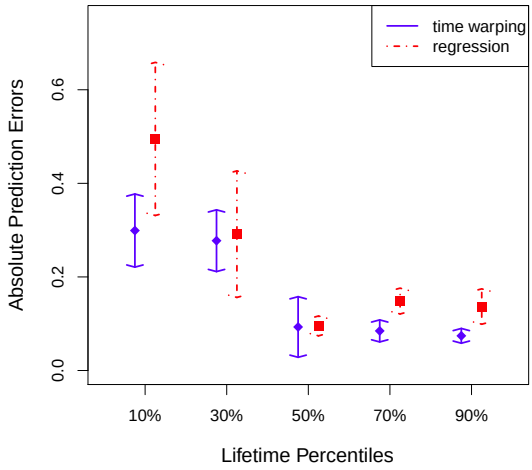


Figure 22: Absolute RLD prediction errors for "time warping" and "exp-brownian"

4.5 Simulation Studies

We conduct three simulation studies for different purposes. In Subsection 4.5.1, we present a detailed simulation study, in which both the estimation and prediction results are reported and discussed. In the simulation study in Subsection 4.5.2, we focus on comparing the prediction performance of our proposed approach with the one proposed in [71]. The third simulation study included in the Appendix B is a sensitivity analysis, in which we assess the prediction performance under departures from the model assumptions and various settings of the parameters.

4.5.1 Simulation Study I

In this subsection, we assess the performance of our proposed functional time warping model by using simulated degradation signals. We repeat the following simulation for 50 times. In each simulation, we randomly generate 70 signals according to the time warping model in (24). The common shape function $Z(\cdot)$ is generated as a linear combination of cubic B-spline basis of the dimension $d_\beta = 5$. The basis coefficient for the common shape function is $\boldsymbol{\beta} = (0, 40, 50, 60, 150)$. The time warping functions $\mu_i(\cdot)$ are also simulated as a linear combination of cubic B-spline basis functions, but of the dimension $d_\phi = 4$. The

basis coefficient for each time warping function follows from a normal distribution defined over the constrained space as given by conditions (20) and (21). More specifically, $\boldsymbol{\phi_i} \sim N(b_0\boldsymbol{\Upsilon}, \boldsymbol{\Omega_\phi}/\lambda_\phi)$ with $b_0 = 100$ and $\lambda_\phi = 100$. The signals are generated with measurement errors and we set $\sigma_\varepsilon = 5$. For each component, we observe one degradation signal that consists of about 100 observations. The failure threshold is specified as $D = 100$.

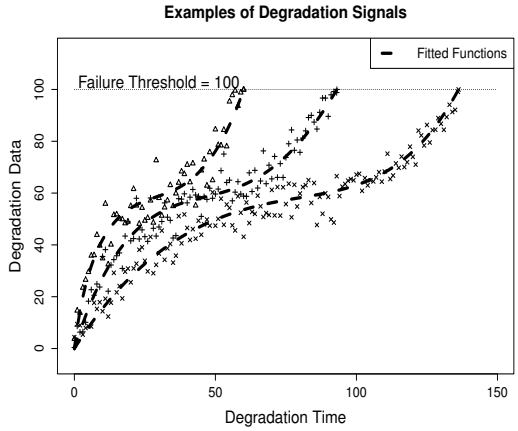


Figure 23: Examples of simulated degradation Signals with their estimated underlying functions.

In each simulation, we randomly select 50 signals as training degradation signals to estimate the model parameters and the rest 20 signals are used as testing signals to evaluate how our model performs in terms of residual life prediction. To determine the tuning parameters d_β and d_ϕ, we compute the Bayes Factor score under different combinations of settings. One example of the Bayes Factor Score (after scaling) versus d_β and d_ϕ is provided in Figure 24. The optimal values of d_β and d_ϕ are determined by maximizing the proposed Bayes Factor score. In this example, they are $(d_\beta, d_\phi) = (7, 4)$, and we will proceed the estimation and prediction procedures using these tuning parameters. Figure 25(a) displays the estimate of the common shape function $Z(\cdot)$ and Figure 25(b) displays the estimate of the underlying degradation trend for one component. In both figures, the solid lines represent the true functions and the dashed lines represent the estimated functions. Both the estimated common shape function and the estimated individual degradation trend function are estimated accurately. One can assess the residuals for the adequacy of the model assumptions. For instance, we can use Pearson chi-square test for assessing the normality

assumptions of residuals or simply visualize the residuals using various graphical techniques. For this simulation study, the residuals do not show any significant pattern and are normally distributed.

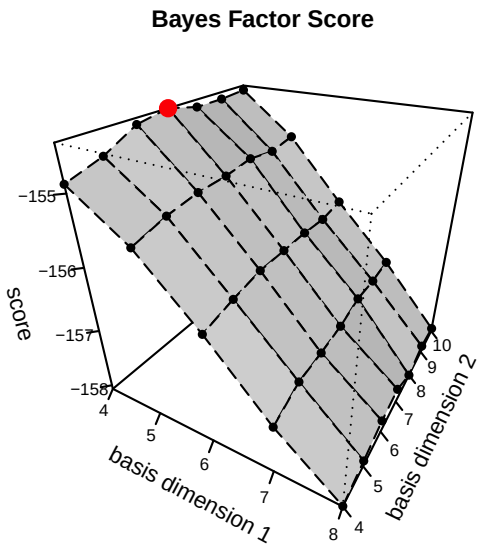
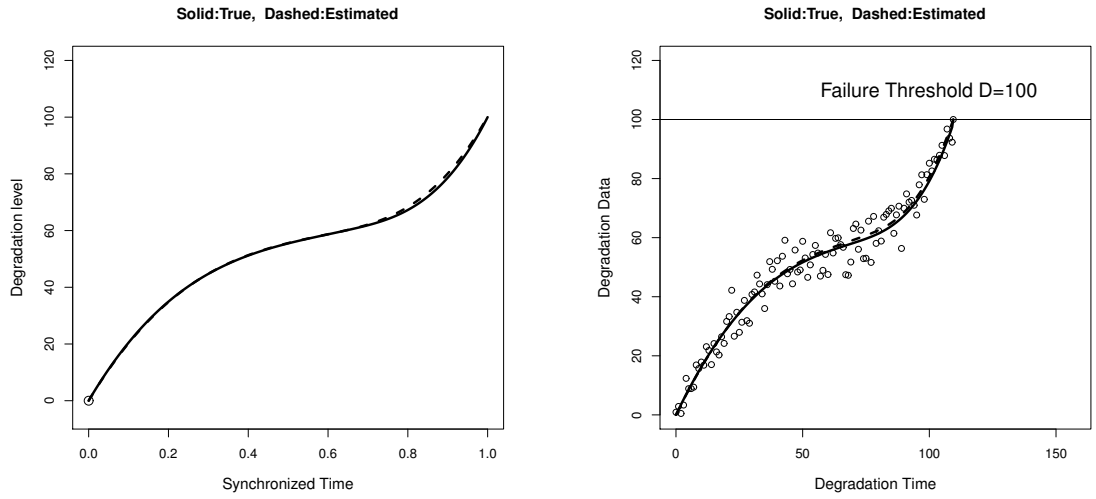


Figure 24: Bayes Factor Score v.s. Tuning Parameters (d_β and d_ϕ)

We update the RLD predictions as more and more observations are collected. This process is illustrated in Figure 26. In this example, we compute the absolute prediction errors described in (26) at varying percentiles of a component's entire life: 20%, 50%, and 80% (80% implies that 80% of the component's life has passed). The degradation observations are displayed in the left figure and the corresponding lifetime predictions are displayed in the right figure.

The prediction results for all test units are summarized and reported in Figure 27, in which the x axis represents the time percentiles and the y axis represents the absolute prediction errors of the residual life. We compute the absolute prediction errors at varying percentiles of a component's entire life: 10%, 30%, . . . , 90%. The median absolute prediction errors corresponding to each percentiles are all less than 10%. This indicates that our proposed model can predict the residual life accurately based on the partial observations of its degradation signal. Another observation is that the median and variance of the absolute



(a) Estimate of $Z(\cdot)$ (b) Estimate of $X_i(\cdot)$

Figure 25: Estimation of Common Shape Function and Individual Underlying Degradation Function

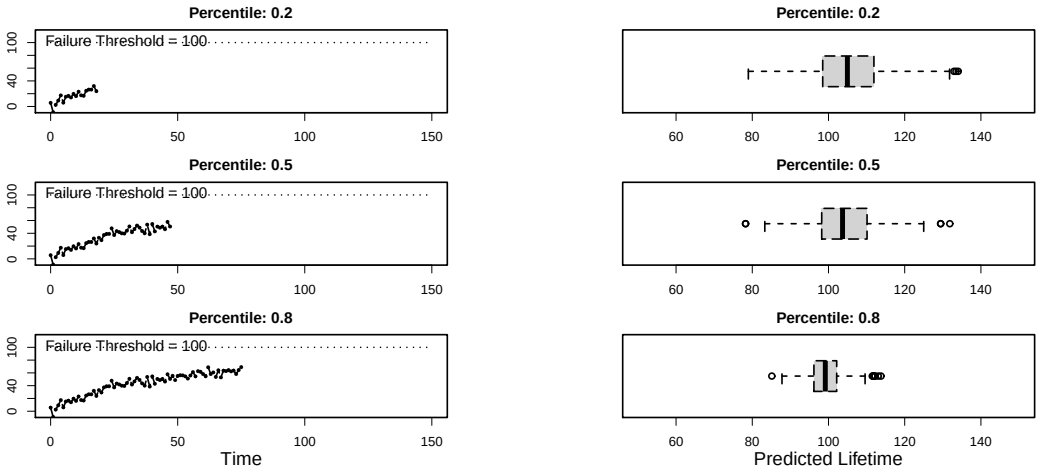
prediction errors decrease with increasing life percentiles. For example, the median absolute prediction error corresponding to the 10th% percentile is around 8% while the median absolute prediction error corresponding to the 90th% percentile is only around 2%. This observation is consistent with our bearing case study, which indicates the benefits of our proposed Bayesian updating framework.

4.5.2 Simulation Study II

In this subsection, we focus on comparing the performance of our proposed approach,"time warping" hereafter, and the approach proposed in [71], "regression" hereafter. We will conduct two simulation studies according to two different model settings:

- Model Setting 1: the degradation signals are simulated based on our proposed model in (24). To simulate $Z(\cdot)$, we set $d_\beta = 5$ and $\boldsymbol{\beta} = (0, 4, 6, 12, 30)$. The time warping functions $\mu_i(\cdot)$ are simulated with $d_\phi = 4$ and $\boldsymbol{\phi_i} \sim N(20\boldsymbol{\Upsilon}, \boldsymbol{\Omega}_{\boldsymbol{\phi}}/5)$ defined over the constrained space as given by conditions (20) and (21). The error standard deviation is $\sigma_\varepsilon = 0.75$. The average number of observations per signal is about 20. We set the failure threshold as $D = 20$.

- Model Setting 2: the degradation signals are generated according to the following



(a) Updating of Signal Observations (b) Updating of Lifetime Predictions

Figure 26: Illustration of the Updating Process

formula: $Y_i(t) = \mu(t) + \xi_i\phi(t) + \epsilon_i(t)$ with $\mu(t) = 100t^2e^{t^2/9}$, $\xi_i \sim Uniform(-60, 60)$, $\phi(t) = t^2$, $0 \leq t \leq 1$, and $\sigma_\varepsilon = 0.1$. The average number of observations per signal is about 15. The failure threshold is set as $D = 2$.

We repeat the following simulation process for 50 times. In each simulation, we randomly generate 35 signals from which we randomly select 25 signals as training signals for estimation and the rest 10 signals are selected as testing signals for validation. We use the error criterion described in (26) to assess the accuracy of the residual life predictions for the two methods, "time warping" and "regression".

The results under the model setting 1 are summarized in Figure 28 (a). In this plot, we show the median and (± 2) standard deviations of the absolute prediction errors at varying lifetime percentiles for "time warping" (blue solid) and "regression" (red dashed). The median absolute prediction errors of "time warping" are consistently smaller than those of "regression" at all lifetime percentiles. This indicates that our proposed method predicts the RLD with higher accuracy. Moreover, the standard deviations of the absolute prediction errors for "time warping" are consistently smaller than those of "regression'. This indicates that our proposed time warping method is more stable in predicting the RLD. Importantly, the difference in the absolute prediction errors between these two approaches is more significant at higher life percentiles. For example, the difference of the median absolute

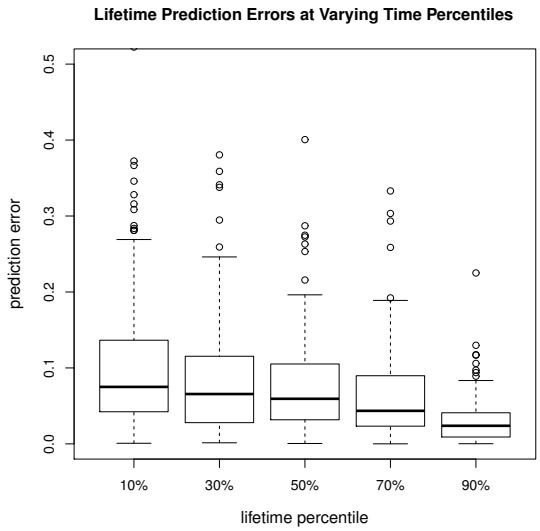
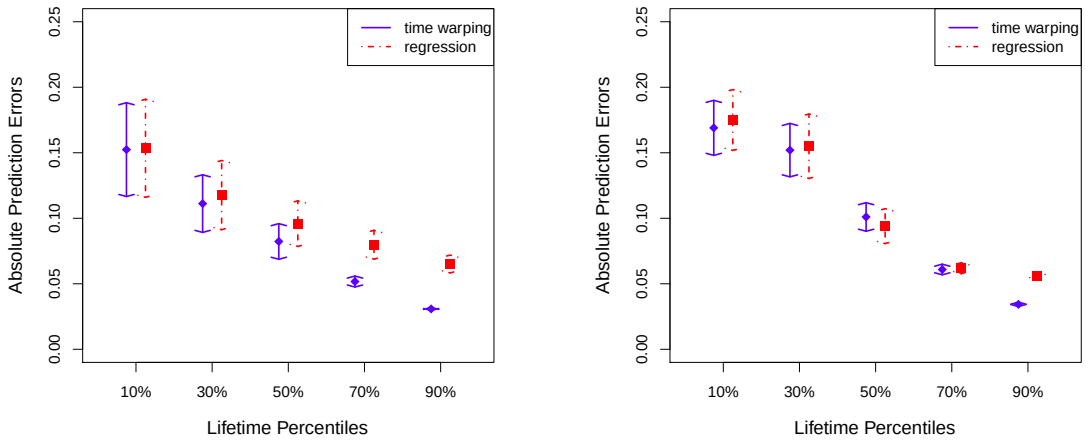


Figure 27: Absolute RLD Prediction Errors

prediction errors is around 0.1% at the 10% percentile, while the difference is around 3.5% at the 90% percentile.

Under the model setting 2, we investigate the robustness in the performance difference between the two approaches to models that not follow the decomposition in (24). We summarize the prediction results in Figure 28 (b). Similar to simulation setting 1, "time warping" has smaller median and standard deviation of absolute prediction errors than those of "regression" although the difference is not as significant as that in the simulation setting 1.



(a) Model Setting 1

(b) Model Setting 2

Figure 28: Absolute RLD prediction errors for "time warping" and "regression"

CHAPTER V

DEGRADATION-BASED RESIDUAL LIFE PREDICTION UNDER DIFFERENT ENVIRONMENTS

In this chapter, we propose a nonparametric framework for modeling degradation signals that may have been collected under different types of environmental conditions. Specifically, we consider two cases, (1) historical degradation signals were collected under known environmental states, and (2) environmental conditions were unknown during the acquisition process for historical data. For the first case, we use a classification algorithm to help identify the environment state of the fielded units. For the second case, an additional step is needed to cluster the historical degradation signals.

The remainder of the chapter is organized as follows. We present the degradation model in Section 5.1. More specifically, we discuss the model framework, model estimation and model prediction in sequence. To assess the model performance, we conduct case studies in Section 5.2, by using simulated degradation signals and degradation signals from rolling bearings. Technical details are deferred in the Appendix C.

5.1 Model Development

5.1.1 Framework

5.1.1.1 Modeling Degradation Signals:

Denote the measurement (or inspection) time by t_{lj}, where $l = 1, \ldots, L$ (L is the number of signals) and $j = 1, \ldots, n_l$ (n_l is the number of observation time points for component l). We assume that the time points are pre-specified within a bounded interval $[0, M]$, where M refers to the maximum experimental time. The degradation amplitude at time t_{lj} is denoted by $S_l(t_{lj})$ and correspondingly, we have $S_l = (S_l(t_{lj}), \cdots, S_l(t_{ln_l}))$.

Note that $S_l(t_{lj})$ may not always be observable. For instance, a component may be shut down or replaced instantaneously after its degradation level reaches the failure threshold. In other words, no further observations can be acquired beyond the failure threshold. This

type of signals are referred to as *truncated degradation signals* in [119]. In these applications, $S_l(t_{lj})$ is observable only if the component has not failed by t_{lj}.

We assume that the underlying degradation process, denoted by $X(\cdot)$, can be represented by a fixed number of basis functions. Based on this assumption, we consider the following degradation model:

$$S_l(t) = X_l(t) + \epsilon_l(t) = B(t)\gamma_l + \epsilon_l(t) \tag{24}$$

where

- $X_l(\cdot)$ represents the underlying degradation process.
- $B(\cdot)$ represents the basis functions of dimension q, defined over the time interval $[0, M]$. For illustrative purposes, we use the cubic B-spline bases in this paper. The choices of basis functions may depend on specific contexts.
- γ_l represents the basis coefficient. It is a vector of dimension q.
- $\epsilon(\cdot)$ represents the error term. We assume that $\epsilon(\cdot)$ is independent and identically distributed at different time points.

5.1.1.2 Modeling Environmental Clusters:

Components may be operated under different types of environmental conditions. In this paper, we assume a component's environmental type is time-stationary, i.e., it does not change over time. Let the environmental type for component l be $Z_l \in \{1, 2, \ldots, K\}$, where K represents the number of environmental types or clusters (in the remaining paper, "environmental types" and "clusters" will be used interchangeably). We make the following distributional assumptions about the clusters:

- Z_l follows a Multinomial distribution with a proportion parameter $\pi = (\pi_1, \ldots, \pi_K)$.
- Conditional on Z_l, the basis coefficient γ follows a normal distribution. The distributional means and variances are different among environmental types. More specifically, $\gamma_{l,k}(t) \equiv \gamma_l | Z_l = k \sim N(\mu_k, \Lambda_k)$, where $\mu_k = (\mu_{k1}, \ldots, \mu_{kq})^T$and Λ_k is a $q \times q$ matrix.

- Conditional on Z_l, the error terms are assumed to follow normal distributions. The variances are different across clusters. In other words, $\epsilon_l(t)|Z_l = k \sim N(0, \sigma_k^2)$.

In summary, we have the following model:

$$\begin{cases} Z_l \sim \text{Multinomial}(\pi_1, \ldots, \pi_K) \\ \gamma_{lk} \equiv \gamma_l|(Z_l = k) \sim N(\mu_k, \Lambda_k) \\ S_l(t) = B(t)\gamma_l + \epsilon_l(t) \\ \epsilon_l(t)|(Z_l = k) \sim N(0, \sigma_k^2) \end{cases} \tag{25}$$

Based on the above formulation, we have: $S_l(t)|Z_l = k \sim N(B(t)\mu_k, B(t)\Lambda_k B(t)^T + \sigma_k^2 I)$.

5.1.2 Estimation

As mentioned earlier, we will consider two possible scenarios, i.e., the cluster membership for the training components may or may not known. In the machine learning context, this corresponds to classification and clustering problems. Let $\mu = (\mu_1, \cdots, \mu_K)$, $\Lambda = (\Lambda_1, \cdots, \Lambda_K)$, $\pi = (\pi_1, \cdots, \pi_K)$, $\sigma = (\sigma_1, \cdots, \sigma_K)$ and $\theta = (\mu, \Lambda, \pi, \sigma)$.

Due to model complexity, it's intractable to maximize the complete data log-likelihood directly with respect to the parameters within θ. To address this problem, we propose an EM algorithm to obtain the maximum likelihood estimates of θ. The estimation procedures are similar for the classification and clustering scenarios, except for an extra step in the clustering case to classify all the training units. Here, we highlight the challenge in estimating the covariance matrix. Other details about the estimation algorithm are provided in the appendix.

5.1.2.1 Estimating the covariance matrix

To allow for more flexibility, we assume in our model that Λ_k are different across clusters. This implies that we need to estimate $\frac{Kq(q+1)}{2}$ parameters for the covariance matrix only. If we do not have a sufficiently large historical data set for training, the covariance matrix estimates will be unstable and inaccurate. To overcome this challenge, we follow the idea of regularized discriminant analysis (RDA) proposed in [29]. More specifically, we regulate the raw covariance matrix estimates in two steps:

Step 1: shrink the individual sample covariance matrix estimate ($\hat{\Lambda}_k$) towards the pooled (population) sample covariance matrix estimate ($\hat{\Lambda}$) with a parameter $0 \leq \lambda \leq 1$.

$$\hat{\Lambda}_k(\lambda) = (1-\lambda)\hat{\Lambda}_k + \lambda\hat{\Lambda}$$

Step 2: shrink $\hat{\Lambda}_k(\lambda)$ towards a multiple of the identity matrix with a parameter $0 \leq \gamma \leq 1$.

$$\hat{\Lambda}_k(\lambda,\gamma) = (1-\gamma)\hat{\Lambda}_k(\lambda) + \gamma\frac{tr(\hat{\Lambda}_k(\lambda))}{p}I$$

In [29], the authors demonstrate through numerous case studies that $\hat{\Lambda}_k(\lambda,\gamma)$ is generally more stable and accurate than the raw covariance matrix estimate $\hat{\Lambda}_k$, especially when the sample size of certain clusters are not large enough. The turning parameters λ and γ can be determined by cross-validation.

5.1.3 Prediction

Given a new fielded component's degradation signal S^* observed up to time t^*, our goal is to predict its residual life RL^*, i.e., the time left for the signal to reach the failure threshold D. In other words, we need to derive the density function $f(RL^*|S^*)$. We approach this in two steps according to the following equation:

$$f(RL^*|S^*) = \int_\gamma f(RL^*|\gamma, S^*)f(\gamma|S^*)d\gamma$$

Step 1: Compute $f(\gamma|S^*)$

$$f(\gamma|S^*) = \sum_{k=1}^{K} f(\gamma|Z^*=k, S^*)P(Z^*=k|S^*)$$

(1) $f(\gamma|Z^*=k, S^*)$ can be computed based on the general Bayesian linear theory. Details can be found in Appendix A;

(2) Since $S^*|Z^*=k \sim N(B\mu_k, B\Lambda_k B^T + \sigma_k^2 I)$, we have:

$$\begin{aligned}
&P(Z^*=k|S^*)\\
&= \frac{f(S^*|Z^*=k)\pi_k}{\sum_{j=1}^{K} f(S^*|Z^*=j)\pi_j}\\
&= \frac{\left|B\Lambda_k B^T+\sigma_k^2 I\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(S^*-B\mu_k)^T(B\Lambda_k B^T+\sigma_k^2 I)^{-1}(S^*-B\mu_k)\right)\pi_k}{\sum_{j=1}^{K}\left|B\Lambda_j B^T+\sigma_j^2 I\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(S^*-B\mu_j)^T(B\Lambda_j B^T+\sigma_j^2 I)^{-1}(S^*-B\mu_j)\right)\pi_j}
\end{aligned}$$

In summary, $\gamma|S^*$ follows a mixture of Gaussian distribution. The means and variances are given in (1) and the mixture probabilities are given in (2).

Step 2: Compute $f(RL^*|S^*)$

For a general path model, such as the one in this paper, it's not possible to derive a closed-form expression of the RLD. Therefore, we resort to a parametric bootstrap methodology ([18, 23]) to generate samples from the distribution of RL^* given the partially observed signal S^*. The details are given as follows:

1. Generate a random sample γ from $f(\gamma|S^*)$ according to the density function given in Step 1;

2. Generate the corresponding simulated signal S_b: $S_b(t) = B(t)\gamma$;

3. Get the lifetime T_b for the generated signal according to the failure time definition: $RL_b = \inf_t\{S_b(t) > D\} - t^*$;

4. If $RL_b > 0$, then proceed to the next step; otherwise, repeat steps $1, 2, 3$ until $RL_b > 0$.

5. Repeat the above steps for N_b times and get N_b values of RL^*: $\mathbf{RL} = (RL_1, RL_2, \ldots, RL_{N_b})$.

RL can then be used for the estimation of any statistics related to RL^*, such as quantiles, prediction intervals and etc.

5.2 Case Studies

In this section, we focus on evaluating the performance of our degradation modeling framework. For this purpose, we conduct two case studies. One is based on simulated degradation signals and another one is based on degradation signals from a rotating machinery experiment.

5.2.1 Simulation Study

5.2.1.1 Generating Signals

In this study, we assume components are from two different clusters, i.e., they are operated under two different environmental types. We first simulate the cluster membership Z_l from a Multinomial distribution with equal proportional parameters, i.e., $Z_l \sim Multinomial(\pi_1 =$

$0.5, \pi_2 = 0.5$). Next, we generate signals for each cluster based on the following model settings:

- In Cluster 1, $S_l(t) = B(t)\gamma_l + \epsilon(t)$, where
  - $B(\cdot)$ represents the cubic B-spline basis with dimension of $q = 5$.
  - $\gamma_l \sim N(\mu_1, \Sigma_1)$, where $\mu_1 = (0, 500, 1500, 2500, 3000)$ and

$$\Sigma_1^{-1} = \Omega_1/5600;\ \Omega_1 = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}_{5\times 5}$$

    Here, μ_1 is selected such that the overall mean degradation trend of this cluster is linear. The underlying degradation process for each unit can still be nonlinear, though. The covariance matrix (Σ_1) of this structure is frequently used as a prior for the basis coefficients under the Bayesian framework [57]. In the frequentest domain, this corresponds to penalized regression splines. More details can be found in [24].
  - $\epsilon(t)|(Z_l = 1) \sim N(0, 80^2)$.
- In Cluster 2, $S_l(t) = \mu(t) + X_l(t) + \epsilon_l(t)$, where
  - $\mu(t) = 4t^2 e^{t/25}$, which represents the overall mean degradation trend for components within this cluster.
  - $X_l(t) = \beta_l t^2$, which is introduced to account for the unit to unit heterogeneity in degradation. Here, $\beta_l \sim N(0, 1.5^2)$.
  - $\epsilon(t)|(Z_l = 2) \sim N(0, 60^2)$.

We note that the signals in Cluster 1 are generated under the general framework proposed in Model 25, while the signals within Cluster 2 are not.

Based on the above model settings, we generate 100 signals (50 for each cluster) for training the degradation model and another 100 signals (50 for each cluster) for evaluating the model performance. All the signals are truncated at the failure threshold, which is $D = 1000$ for both clusters. We will evaluate the performance of our methodology under complete as well as sparse scenarios. For a complete signal, the measurement time points are pre-set at an equally spaced grid $c_0, ..., c_{80}$ on $[0, 20]$ with $c_0 = 0, c_{80} = 20$. For a sparse signal, we uniformly sample 12 time points from $c_0, ..., c_{80}$. Note that these time points are pre-specified; due to truncation, degradation signals at these time points are not always observable. In this simulation, we have around 40 observation per complete signal and only around 6 observations per sparse signal. Examples of the generated complete and sparse degradation signals can be found in Figure 29(a) and Figure 29(b), respectively. In both plots, the black signals are from Cluster 1 and the grey signals are from Cluster 2.

5.2.1.2 Estimation

We estimate the mean degradation trend for each cluster based on complete or sparse signals. The results are shown in Figure 30(a) and Figure 30(b), respectively. In both plots, the solid, dashed and double dashed lines represent the true mean trend, the estimated trend in the clustering scenario and the estimated trend in the classification scenario, respectively. From Figure 30, we see that the estimated mean degradation trend for both classification and clustering scenarios are very close to the true one (except for a small departure in the second cluster under the sparse case, due to limited data available in that region). This indicates that the mean functions are estimated very accurately for both classification and clustering scenarios and under both complete and sparse cases.

In the clustering scenario, we are also interested in whether the training signals are clustered accurately, i.e., whether the resulting clusters are consistent with the actual clusters. To this purpose, we compute the Rand index between them. The Rand index measures the percentage of pairs of components on which two clusterings, denoted by X_1 and X_2, agree or disagree. Assume that there are N components available for clustering, then there are $N(N-1)/2$ distinct pairs. Denote the number of pairs that are in the same cluster in both

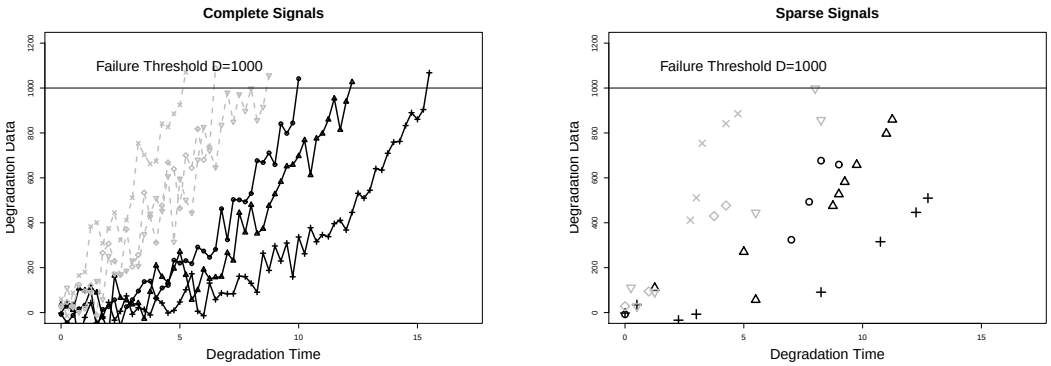


Figure 29: Examples of complete (left plot) and sparse (right plot) degradation signals

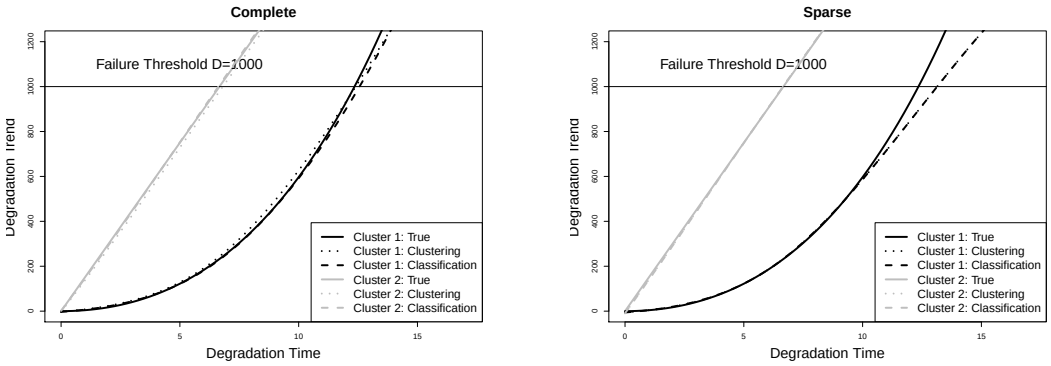


Figure 30: Estimated (true, classification, clustering) mean function for the complete and sparse scenarios, respectively

X_1 and X_2 by n_{00} and the number of pairs that are in different clusters in both X_1 and X_2 by n_{11}, then the Rand index of X_1 and X_2 is defined as:

$$Rand(X_1, X_2) = \frac{n_{00} + n_{11}}{N(N-1)/2}$$

$Rand(X_1, X_2)$ ranges from 0, when there are no pairs classified in the same way under X_1 and X_2, to 1, when X_1 and X_2 are identical. For our purpose, X_1 represents the true cluster membership and X_2 represents our proposed clustering method. In this simulation, the Rand index is 0.99 under both complete and sparse scenarios. This strongly indicates that our clustering method performs well under both scenarios.

5.2.1.3 Prediction

Our next step is to evaluate the performance of our model in terms of residual life prediction. To assess the prediction accuracy, we utilize the following error criteria:

$$\text{Prediction Error} = \frac{|\text{Predicted Life - Actual Life}|}{\text{Actual Life}} \tag{26}$$

For each testing component, we compute the prediction errors at the following percentiles of its entire life: 10%, 30%, ..., 90% (90% implies that 90% of the component's life has passed). The results based on complete and sparse signals are illustrated in Figure 31 and Figure 32, respectively. In both figures, the left and middle plots are for the classification and clustering cases. For comparative purposes, we also use a benchmark method, which is based on our proposed framework (including the estimation and prediction approaches). However, this benchmark method assumes all the components are from the same population and therefore, it does no classifications or clusterings. We refer to this benchmark method as "no clustering". The results of "no clustering" are reported in the right plots of Figures 31 and 32. For ease of comparisons, we also summarize the results in Table 3, which gives the mean and variance of prediction errors for different methods based on complete and sparse degradation signals.

One consistent observation from Figure 31, Figure 32 and Table 3 is that both the mean and variance of the prediction errors decrease as the lifetime percentile increases. The improvement is due to our proposed Bayesian updating framework. The updating

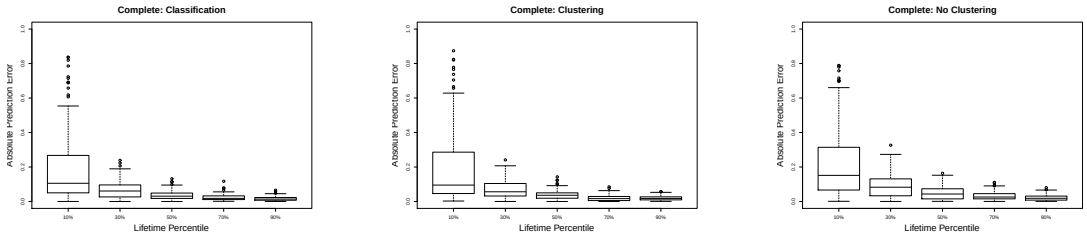


Figure 31: Absolute prediction errors for "classification", "clustering" and "no clustering" based on complete degradation signals

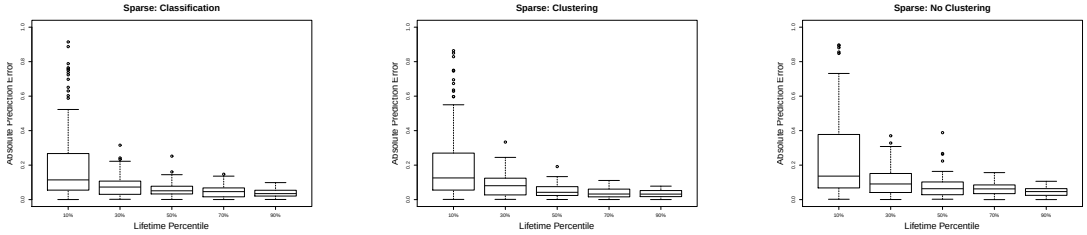


Figure 32: Absolute prediction errors for "classification", "clustering" and "no clustering" based on sparse degradation signals

Table 3: Lifetime prediction results of "classification", "clustering" and "no clustering" under complete and sparse scenarios

Lifetime Percentiles	10%	30%	50%	70%	90%
Complete: classification	0.2032	0.0679	0.0362	0.0230	0.0170
Complete: clustering	0.2026	0.0686	0.0387	0.0203	0.0201
Complete: no clustering	0.2414	0.0885	0.0491	0.0338	0.0224
Sparse: classification	0.2131	0.0807	0.0597	0.0410	0.0371
Sparse: clustering	0.2136	0.0863	0.0519	0.0407	0.0347
Sparse: no clustering	0.2553	0.1070	0.0727	0.0624	0.0449

process allows for updating the parameter distributions using real-time data from a fielded component, which further results in a better prediction of the components' RLD.

Another observation is that the prediction results are very similar for the classification and clustering scenarios. This again demonstrates that, at least in this simulation, our proposed clustering algorithm can accurately classify signals of similar patters into the same group and separate signals of distinct patterns into different groups. Furthermore, the benchmark method provides less accurate predictions of the residual life of fielded components. This is because the assumption that all components are from the same population does not hold in this simulation, and therefore, the prior derived based on this assumption is inaccurate. The difference of performance between our methods and the benchmark method "no clustering" is more remarkable at smaller life percentiles, when the prior plays a relatively more important role in the RLD predictions.

5.2.2 Bearing Case Study

Bearings play an important role in a wide range of engineering applications, particularly in rotating machinery. Failures of bearings can lead to unexpected shutdown or failure of the entire engineering system. In this study, we conduct an experiment to monitor the degradation processes of rolling bearings. Each bearing is operated under one of the following two rotational levels: 2600 and 2200 rpm (rpm is shorted for "revolutions per minute"). The sample size of each cluster is 18 and 16, respectively. For all these bearings, we collect vibration-based degradation signals up to their failures. The failure threshold is pre-specified as $D = 0.02$ Vrms (Vrms is shorted for "vibrational root mean square"). Examples of the resulting degradation signals are provided in Figure 33. In this figure, the grey lines represent the degradation signals from cluster 1 (2600 rpm) and the black lines represent those from cluster 2 (2200 rpm).

To evaluate the performance of our proposed degradation model, we repeat the following study for 50 times. Each time we randomly select 5 signals from each cluster as the testing signals. For these testing signals, we assume that their cluster membership, or rotational speeds, are unknown and need to be predicted. These testing signals will be used to assess

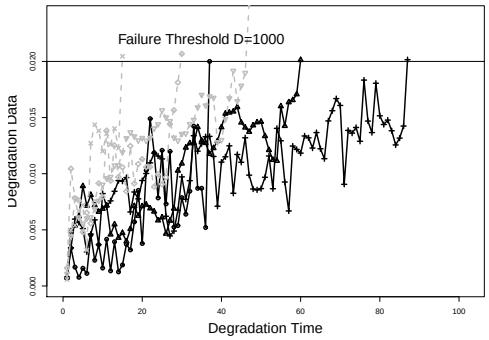


Figure 33: Examples of bearing degradation signals

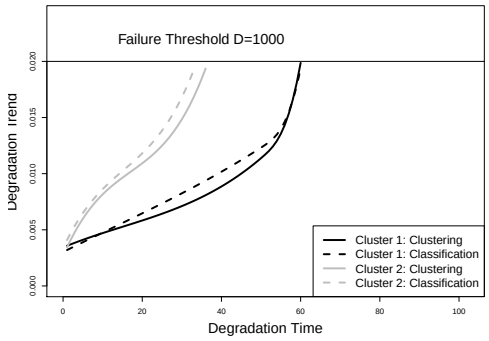


Figure 34: Estimated mean degradation trend under the classification and clustering scenarios

the prediction performance. The rest of 24 degradation signals (13 of them are from cluster 1 and the rest 11 signals are from cluster 2) form a historical data set and they are used to train the proposed degradation model and estimate the parameters therein. Depending on the scenario we are interested in, whether it's "classification" or "clustering", the cluster membership for the training components may or may not be known. In Figure 34, we show the estimated mean degradation trend (up to the failure threshold) for both clusters. Apparently, the degradation processes in cluster 1 with a rotational speed of 2600rpm is relatively faster than those from cluster 2 with a rotational speed of 2200rpm. Another observation is that, the estimated mean degradation trend under the classification and clustering scenarios are close to each other.

To mimic and illustrate the real-time updating process, we also assess the prediction performance progressively. More specifically, for each test bearing, we predict its residual life by using the partially observed signal at the following percentiles of its lifetime: 10%,

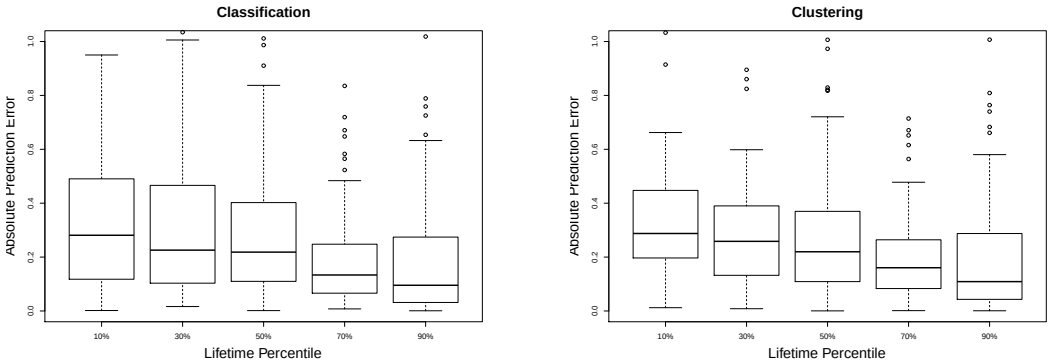


Figure 35: Absolute prediction errors under the classification and clustering scenarios

30%, $\cdots$, 90%. As the percentile gets larger, we have more degradation observations available, and therefore, we expect to see more accurate and more precise predictions of the RLD. This is demonstrated in the boxplots of Figure 35. In Figure 35, we consider both the classification and clustering scenarios. Among these boxplots, the x-axis represents the lifetime percentiles and the y-axis records the absolute prediction errors decried in Equation 26. We observe that both the median and variance of the prediction errors decrease as the lifetime percentile increases, and this is consistent with our previous expectation. Another observation from Figure 35 is that, the prediction performance for the classification and clustering scenarios are very similar. This, once again, demonstrates that our proposed clustering algorithm can classify degradation signals into reasonable clusters.

CHAPTER VI

FUTURE WORK

6.1 Future Work

In Chapter 5, we extend our previous work to the case when degradation signals are observed from different types of environmental conditions. The major assumptions of this work are as follows:

1. The environmental conditions can be categorized into a small set of discrete types;

2. The environmental conditions are time-invariant.

In the first assumption, we require the number of clusters is small such that we have enough signals for each cluster and that we have accurate and stable estimation of model parameters pertaining to each cluster. Intuitively speaking, the clustering scenario is relatively more robust to this assumption since the number of clusters is determined by cross validation; a large value of the number of clusters will not be selected due to relatively larger estimation and prediction errors. In the classification scenario, however, the number of clusters is pre-fixed. Therefore, it could be possible that the number of clusters is big and there exist certain clusters with very few signals. In the extreme case, the number of clusters can be equal to the number of signals. One remedy for this is to incorporate an agglomerative clustering algorithm to combine those smaller groups of signals.

The environmental conditions can be time-variant in two different types. One case is that the environmental conditions vary over time continuously and smoothly. In this case, we can complement the framework developed in Chapter 5 with an extra (linear, nonlinear or nonparametric) regression model, with the environmental conditions Z_l and the basis coefficients γ_l being covariates and the response variable, respectively. Another case is that the environmental conditions are still discrete, but they may change at some deterministic or random time epochs [9, 35]. At these transitional epochs, the observed degradation signals

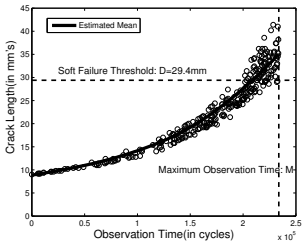
may be subject to sudden shocks, and the rate at which the degradation progresses may also change after these epochs. A further extension of the present framework to incorporate such time-varying environmental conditions will be of interest in our future research.

APPENDIX A

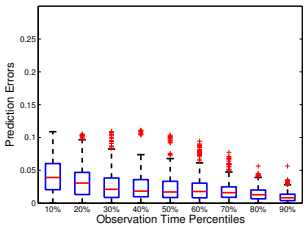
SUPPLEMENTARY MATERIALS OF CHAPTER II

A.1 Additional Results of Crack Growth Data Study

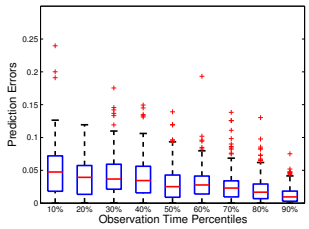
This section provides prediction results for the real data used in the accompanying paper with a different soft failure threshold $D = 29.4$ mm. The prediction results are illustrated in Figure 36. The plots are organized in the same order as those in Section 6. Based on these plots, our findings are consistent with the ones reported in Section 6 in the main paper.



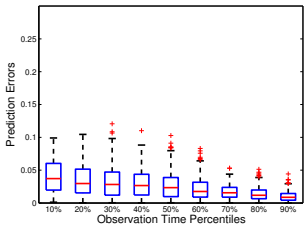
(a) Examples of Degradation Signals



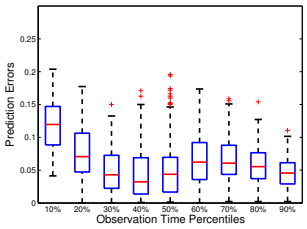
(b) Complete: Non-parametric Model



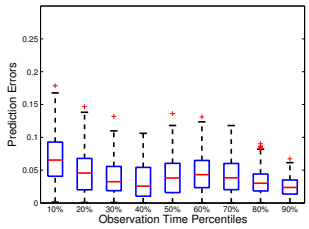
(c) Fragmented: Non-parametric Model



(d) Sparse: Non-parametric Model



(e) Sparse: Log-linear

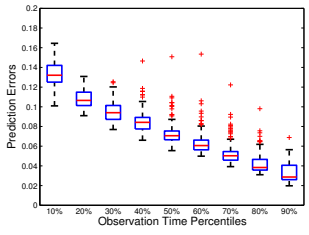


(f) Sparse: Log-log-linear

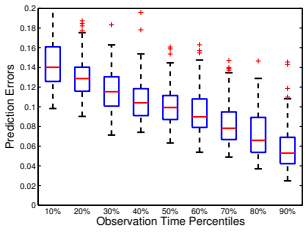
Figure 36: The prediction error of residual life prediction for the crack growth data.

A.2 Additional Results of Simulation Model 2

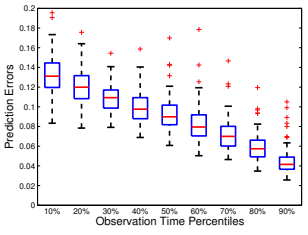
In the main chapter we reported only the prediction errors comparing the non-parametric and parametric models under non-uniform sampling. In the supplemental plots in Figure 37, we provide additional results comparing the prediction errors for incomplete (sparse and fragmented) to complete signals and comparing the prediction errors under departures from the underlying model (e.g. t-distribution for the scores and t-distribution for the errors).



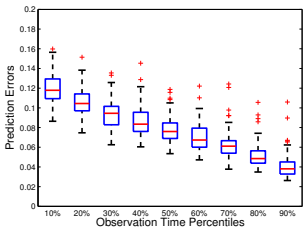
(a) Complete: Non-parametric Model



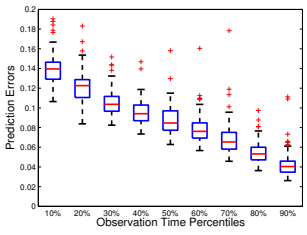
(b) Fragmented: Non-parametric Model



(c) Sparse: Non-parametric Model: Gamma Distribution for the Scores



(d) Sparse: Non-parametric Model: t-Distribution for the Scores



(e) Sparse: Non-parametric Model: t-Distribution for the Errors

Figure 37: The prediction error of the residual life estimate for Model 2.

APPENDIX B

SUPPLEMENTARY MATERIALS OF CHAPTER IV

B.1 Model Selection

The specification of the two tuning parameters, d_β and d_ϕ, is essential in our modeling framework. This can be addressed using model selection approaches. For instance, one can implement cross-validation and select (d_β, d_ϕ) to minimize a predefined measure of the prediction errors. Another alternative is to determine (d_β, d_ϕ) to maximize a penalized likelihood, such as AIC and BIC. In this paper, we determine (d_β, d_ϕ) according to the Bayes Factor as it is consistent with our Bayesian estimation framework.

Denote the data by D, and the candidate models by M_k, for $k = 1, \ldots, K$, where K represents the number of values of (d_β, d_ϕ). The Bayes factor associated with two competing models M_j and M_k is defined as follows:

$$B_{jk} = \frac{f(D|M_j)}{f(D|M_k)}$$

The model M_j is favoured if B_{jk} is bigger than 1. To compute the density $f(D|M_k)$, we need to integrate over the parameter space:

$$f(D|M_k) = \int f(D|\theta_k, M_k)\pi(\theta_k|M_k)d\theta_k$$

where θ_k is the parameter under M_k, $\pi(\theta_k|M_k)$ is its prior density, and $f(D|\theta_k, M_k)$ is the probability density of D given the value of θ_k.

Various estimators of $f(D|M_k)$ are discussed in Kass and Raftery (1995). In this paper, we estimate $f(D|M_k)$ using the harmonic mean of the likelihood values:

$$\hat{f}_1(D|M_k) = \left\{ \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} f(D|\theta^{(i)}, M_k)^{-1} \right\}^{-1}$$

where $\theta^{(i)} : i = 1, \ldots, N_{mc}$ is a sample from the posterior distribution. It is demonstrated that $\hat{f}_1(D|M_k)$ converges almost surely to the correct value, $f(D|M_k)$, as $N_{mc} \to \infty$. In

our paper, we can compute $\hat{f}_1(D|M_k)$, since the posterior samples of $\theta^{(i)}$ are generated in the MCMC estimation step. After $\hat{f}_1(D|M_k)$ is computed for $j = 1, \ldots, K$, one can choose (d_β, d_ϕ) to maximize $\hat{f}_1(D|M_k)$.

B.2 Estimation and Prediction

B.2.1 Estimation

In this part, we use a Gibbs-within-Metropolis algorithm to estimate model parameters. Let N_{MC} represent the number of MCMC runs, then we follow the following steps until $k = N_{MC}$:

- *Step 0*: set initial values $\boldsymbol{\theta^{(0)}}$ for all parameters and set $k = 1$. Here k represents the kth run of the MCMC procedure.
- *Step 1*: use the Gibbs sampling algorithm to get estimates for all parameters, except $\boldsymbol{\phi}$. Their respective closed-form full conditionals are given in the main paper. We denote the resulting estimates by $\hat{\boldsymbol{\theta}}^{(\boldsymbol{k})}_{-\boldsymbol{\phi}}$.
- *Step 2*: get estimates of $\boldsymbol{\phi^{(k)}} = (\phi_1^{(k)}, \cdots, \phi_I^{(k)})$ as follows:
  - *Step 2.1*: set $\phi_{i,1}^{(k)} = 0$ and $\phi_{i,d_\phi}^{(k)} = T_{i,o}$. Here we use $T_{i,o}$, the observed failure time of component i, as an approximation of the underlying failure time $T_{i,u}$. This is where we enforce the image assumption to hold.
  - *Step 2.2*: set $j = 2$, run the following sub-steps until $j = d_\phi - 1$:
    - *Step 2.2.1*: generate a random number $\phi_{i,j}^{(k)}$ uniformly from the support $[\phi_{i,j-1}^{(k)}, \phi_{i,j+1}^{(k-1)}]$ This is where we enforce the monotonicity assumption to hold.
    - *Step 2.2.2*: use the Metropolis Hastings algorithm to determine whether $\phi_{i,j}^{(k)}$ should be accepted. If accepted, then $\hat{\phi}_{i,j}^{(k)} = \phi_{i,j}^{(k)}$, otherwise, $\hat{\phi}_{i,j}^{(k)} = \hat{\phi}_{i,j}^{(k-1)}$
    - *Step 2.2.3*: set $j = j + 1$.
- *Step 3*: set $k = k + 1$.

B.2.2 Prediction

In this part, we only need to predict the parameter $\boldsymbol{\phi}^*$. Let N^*_{MC} represent the number of MCMC runs, then we follow the following steps until $k = N^*_{MC}$:

- *Step 0*: set initial values $\boldsymbol{\phi}^{(\mathbf{0*})}$ for $\boldsymbol{\phi}^*$ and set $k = 1$.
- *Step 1*: get estimates of $\boldsymbol{\phi}^{(\boldsymbol{k*})}$ as follows:
  - *Step 1.1*: set $\phi_1^{(k*)} = 0$. This is where we enforce the image assumption to hold.
  - *Step 1.2*: set $j = 2$, run the following sub-steps until $j = d_\phi$:
    - *Step 1.2.1*: generate a random number $\phi_j^{(k*)}$ uniformly from the support $[\phi_{j-1}^{(k*)}, \phi_{j+1}^{(k-1)*}]$ (Note that if $j = d_\phi$, then the support would be $[\phi_{j-1}^{(k*)}, M]$, where M represents the maximum possible failure time). This is where we enforce the monotonicity assumption to hold.
    - *Step 1.2.2*: use the Metropolis Hastings algorithm to determine whether $\phi_j^{(k*)}$ should be accepted. If accepted, then $\hat{\phi}_j^{(k*)} = \phi_j^{(k*)}$, otherwise, $\hat{\phi}_j^{(k*)} = \hat{\phi}_j^{(k-1)*}$
    - *Step 1.2.3*: set $j = j + 1$.
- *Step 2*: set $k = k + 1$.

B.3 Simulation Study III

In this additional simulation study, we perform a sensitivity analysis. We generate degradation signals from the same model as the one used in the simulation study I, but with assumption violations or under different parameter settings.

B.3.1 Sensitivity to Assumption Violations

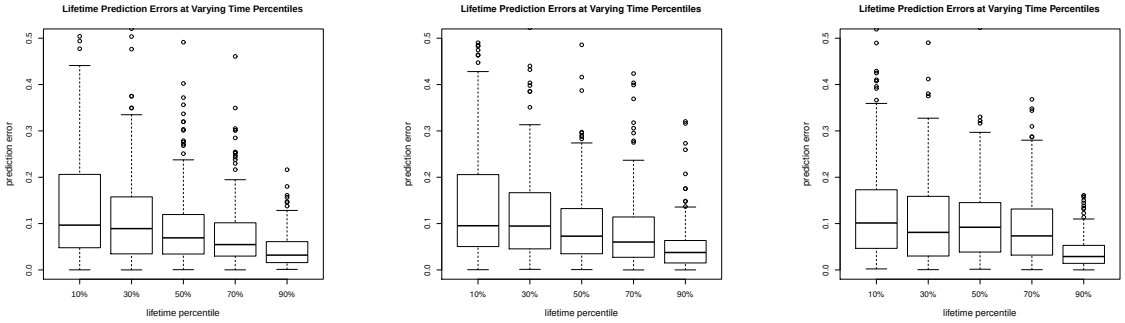
Our modeling framework is based on the following two assumptions:

- ϕ_i follows a normal distribution, i.e., $\phi_i \sim Nor(b_0\Upsilon, \Sigma_\phi)$. For the sensitivity analysis, we instead generate degradation signals from the following two distributions:

– a mixture of gaussian distributions: $p(\phi_i) = \frac{1}{2}Nor(b_0^1\Upsilon, \Sigma_\phi^1) + \frac{1}{2}Nor(b_0^2\Upsilon, \Sigma_\phi^2)$. We input the values of b_0^1, b_0^2, Σ_ϕ^1 and Σ_ϕ^2 such that the simulated degradation signals have approximately the same mean and standard deviation as those of the signals simulated in the simulation study I. The prediction results are in Figure 38(a).

– a non-cental t distribution: $\phi_i \sim t(\mu_t, \Sigma_t, df = 3)$, where μ_t and Σ_t are selected such that the simulated degradation signals have approximately the same mean and standard deviation as those of the signals simulated in the simulation study I. The prediction results are in Figure 38(b).

- c_i and a_i are constant among different units ensuring that the image assumption holds. To test the sensitivity of our framework to this assumption, c_i and a_i are assumed to be random in this study. Specifically, c_i and a_i are simulated from uniform distributions: $c_i \sim Uniform[-2.5, 2.5]$ and $a_i \sim Uniform[0.975, 1.025]$. The prediction results are in Figure 38(c).



(a) mixture gaussian (b) noncental t (c) random c and a

Figure 38: Residual Life Prediction Errors under Assumption Departures

The prediction results in Figure 38 of this supplemental material and in Figure 27 of the main manuscript are similar. Therefore, we validate that our modeling framework is not sensitive to its main assumptions.

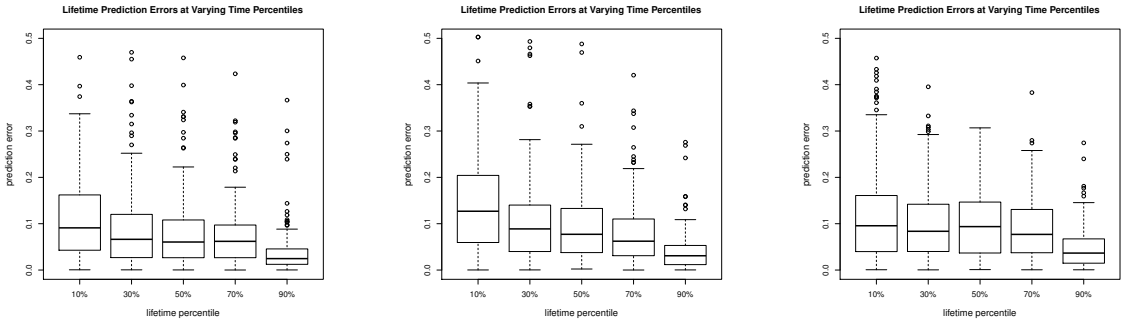
B.3.2 Sensitivity to Parameter Settings

To test the sensitivity of our model to its parameter settings, we repeat the simulation study I, but under different settings of N, d, and σ_ϵ.

- Setting 1: $N = 20$ (In simulation study I, $N = 50$) and all the other settings are the same as those in the simulation study I. The prediction results are in Figure 39(a).

- Setting 2: $d = 5$ (In simulation study I, $d = 1$) and all the other settings are the same as those in the simulation study I. Under this setting, we have around around 20 observations per signal. The prediction results are in Figure 39(b).

- Setting 3: $\sigma_\epsilon = 7.5$ and all the other settings are the same as those in the simulation study I. The prediction results are in Figure 39(c).



(a) Sub-study 1: N=20 (b) Sub-study 2: d=5 (c) Sub-study 3: $\sigma_\epsilon = 7.5$

Figure 39: Residual Life Prediction Errors under More Difficult Parameter Settings

The prediction results in Figure 39 in this supplemental material and in Figure 27 in the main manuscript are similar. When we compare the prediction performance across various settings, we conclude that the magnitude of the measurement error variance is probably more critical than the number of training signals or the number of observations per signal.

APPENDIX C

SUPPLEMENTARY MATERIALS OF CHAPTER V

C.1 An Introductory Lemma

This appendix is devoted to presenting a lemma based on Bayesian linear theory. This lemma will be used in our estimation and prediction algorithms.

According to Lemma 1 in [62], given a model $W = P\xi + \epsilon$ with the assumption that the error $\epsilon \sim N(0, \sigma^2 I)$ and the coefficient ξ has a prior distribution specified by $\xi \sim N(\theta, \Lambda)$, the posterior distribution of ξ is

$$\xi | W, \theta, \Lambda \ \sim \ N(Cd, C)$$

with

$$\begin{aligned} C &= \left(\frac{1}{\sigma^2} P'P + \Lambda^{-1} \right)^{-1} \\ d &= \frac{1}{\sigma^2} P'W + \Lambda^{-1}\theta \end{aligned}$$

C.2 EM Algorithm for Model Estimation

This appendix is devoted to presenting our estimation algorithm. In this algorithm, we focus on the classification scenario, i.e., Z_l is observable for $l = 1, \cdots, L$.

For component l, let S_l^o be the part of signal that is indeed observed and S_l^c be the set of missing signal, then $S_l = (S_l^c, S_l^o)$. Our complete data is composed of the observed part: (1) the observed degradation signals $S^o = (S_1^o, \ldots, S_L^o)$; (2) the cluster membership $Z = (Z_1, \ldots, Z_L)$; and the unobserved/missing part: (1) the missing degradation signals $S^c = (S_1^c, \ldots, S_L^c)$, (2) the basis coefficients $\gamma = (\gamma_1, \ldots, \gamma_L)$. Let $S = (S_1, \cdots, S_L)$, our complete data is then composed of (S, γ, Z). The complete-data likelihood L_C is:

$$L_C(S, Z, \gamma) = f(S|Z, \gamma) f(\gamma|Z) f(Z) = \prod_{l=1}^{L} \left[f(S_l|Z_l, \gamma_l) f(\gamma_l|Z_l) f(Z_l) \right]$$

The densities in the above equation are given as follows:

$$
\begin{aligned}
&f(Z_l;\pi) = \sum_{k=1}^{K} z_{lk}\pi_k \\
&f(\gamma_l|Z_l;\mu_k,\Lambda_k) = \sum_{k=1}^{K} z_{lk}\left[\frac{1}{(2\pi)^{q/2}|\Lambda_k|^{1/2}}\exp\left\{-\frac{1}{2}(\gamma_l-\mu_k)^T\Lambda_k^{-1}(\gamma_l-\mu_k)\right\}\right] \\
&f(S_l|Z_l,\gamma_l;\sigma^2) = \sum_{k=1}^{K} z_{lk}\left[\frac{1}{(2\pi\sigma^2)^{-n_l/2}}\exp\left\{-\frac{1}{2\sigma^2}(S_l-B\gamma_l)^T(S_l-B\gamma_l)\right\}\right]
\end{aligned}
$$

Therefore, the complete-data log-likelihood l_C (up to a constant) is:

$$
\begin{aligned}
&l_C(S,\gamma,Z) = \log(L_C(S,\gamma,Z)) \\
&= \sum_{l=1}^{L}\left[\log f(S_l|Z_l,\gamma_l) + \log f(\gamma_l|Z_l) + \log f(Z_l)\right] \\
&= -\tfrac{1}{2}\sum_{l=1}^{L}\sum_{k=1}^{K} z_{lk}\left[n_l\log\sigma^2 + \tfrac{1}{\sigma^2}\|S_l-B\gamma_l\|^2\right] \\
&-\tfrac{1}{2}\sum_{l=1}^{L}\sum_{k=1}^{K} z_{lk}\left[\log|\Lambda_k| + (\gamma_l-\mu_k)^T\Lambda_k^{-1}(\gamma_l-\mu_k)\right] + \sum_{l=1}^{L}\sum_{k=1}^{K} z_{lk}\log(\pi_k) \\
&= A.1 + A.2 + A.3
\end{aligned}
$$

In the above expression, n_l represents the number of observations for the lth signal, and

$$
z_{lk} = \begin{cases} 1 & \text{if } Z_l = k \\ 0 & \text{if } Z_l \neq k \end{cases}
$$

C.2.1 E-step:

In this step, we need to compute the distribution of the missing data (γ_l, S_l) given the observed data (Z_l, S_l^o):

$$
f(\gamma_l, S_l|Z_l, S_l^o) = f(\gamma_l|S_l, Z_l, S_l^o)f(S_l|Z_l, S_l^o) = f(\gamma_l|S_l, Z_l)f(S_l|Z_l, S_l^o)
$$

Step E.1: Compute $f(\gamma_l|S_l, Z_l)$

According to the Bayes Linear Theory (see Appendix A.1), we first have:

$$
\gamma_l|(Z_l = k, S_l) \sim N((\Lambda_k^{-1} + \frac{1}{\sigma^2}B^TB)^{-1}(\frac{1}{\sigma^2}B^TS_l + \Lambda_k^{-1}\mu_k), (\Lambda_k^{-1} + \frac{1}{\sigma^2}B^TB)^{-1}))
$$

Step E.2: Compute $f(S_l|Z_l, S_l^o)$

Here we provide a monte carlo procedure to generate random samples from this distribution. For S_l, since the part of S_l^o is observed, the sampling procedure is reduced to

generating random samples of S_l^c according to $f(S_l^c|Z_l, S_l^o, C(S_l^c) = 1)$, where $C(S_l^c) = 1$ means that S_l^c is truncated.

According to the Model 25, we have $S_l|Z_l = k \sim N(\widehat{\mu}_l, \widehat{\Sigma}_l)$ where $\widehat{\mu}_l = B\mu_k, \widehat{\Sigma}_l = B\Lambda_k B^T + \sigma^2 I$. We partition $\widehat{\mu}_l$ and $\widehat{\Sigma}_l$ as follows (according to whether the observations are truncated) :

$$\widehat{\mu}_l = (\mu_l^o, \mu_l^c), \widehat{\Sigma}_l = \begin{bmatrix} \Sigma_l^o & \Sigma_l^{oc} \\ \Sigma_l^{co} & \Sigma_l^c \end{bmatrix}$$

Using the property of multivariate normal distribution, we have the conditional distribution of S_l^c given S_l^o is:

$$S_l^c|S_l^o, Z_l = k \sim N(\mu_l^{c|o}, \Sigma_l^{c|o})$$

where

$$\mu_l^{c|o} = \mu_l^c + \Sigma_l^{co}(\Sigma_l^o)^{-1}\left[S_l^o - \mu_l^o\right], \Sigma_l^{c|o} = \Sigma_l^c - \Sigma_l^{co}(\Sigma_l^o)^{-1}\Sigma_l^{oc}$$

Since S_l^c is truncated, $S_l^c|Z_l, S_l^o, C(S_l^c) = 1$ follows a truncated normal distribution, with mean $\mu_l^{c|o}$, and variance $\Sigma_l^{c|o}$, and truncation point at the failure threshold D.

Step E.3: Compute $E(\gamma_l|Z_l, S_l^o)$, $E(\gamma_l\gamma_l^T|Z_l, S_l^o)$, and $E(\gamma_l^T\gamma_l|Z_l, S_l^o)$

These quantities are computed as they will be used in the maximization step. Assume that after Step E.2, we have R samples of S_l: $(S_l{}^{(1)}, \ldots S_l{}^{(R)})$, then:

$E(\gamma_l|Z_l, S_l^o) = \int_{S_l|Z_l,S_l^o} E(\gamma_l|Z_l, S_l) f(S_l|Z_l, S_l^o) dS_l \approx \frac{1}{R}\sum_{r=1}^{R} E(\gamma_l|Z_l, S_l{}^{(r)})$

$E(\gamma_l\gamma_l^T|Z_l, S_l^o) = \int_{S_l|Z_l,S_l^o} E(\gamma_l\gamma_l^T|Z_l, S_l) f(S_l|Z_l, S_l^o) dS_l$

$\approx \frac{1}{R}\sum_{r=1}^{R} E(\gamma_l\gamma_l^T|Z_l, S_l{}^{(r)}) = \frac{1}{R}\sum_{r=1}^{R}\left[E(\gamma_l|Z_l, S_l{}^{(r)})E(\gamma_l^T|Z_l, S_l{}^{(r)}) + Var(\gamma_l|Z_l, S_l{}^{(r)})\right]$

$E(\gamma_l^T\gamma_l|Z_l, S_l^o) = \int_{S_l|Z_l,S_l^o} E(\gamma_l^T\gamma_l|Z_l, S_l) f(S_l|Z_l, S_l^o) dS_l$

$\approx \frac{1}{R}\sum_{r=1}^{R} E(\gamma_l^T\gamma_l|Z_l, S_l{}^{(r)}) = \frac{1}{R}\sum_{r=1}^{R}\left[E(\gamma_l^T|Z_l, S_l{}^{(r)})E(\gamma_l|Z_l, S_l{}^{(r)}) + trace(Var(\gamma_l|Z_l, S_l{}^{(r)}))\right]$

where $E(\gamma_l|Z_l, S_l{}^{(r)})$ and $Var(\gamma_l|Z_l, S_l{}^{(r)})$ are computed according to step E.1. The last two equations are derived since we have:

$$E(\vec{\alpha}\vec{\alpha}^T) = E(\vec{\alpha})E(\vec{\alpha}^T) + Var(\vec{\alpha})$$
$$E(\vec{\alpha}^T\vec{\alpha}) = E(\vec{\alpha}^T)E(\vec{\alpha}) + trace(Var(\vec{\alpha}))$$

C.2.1.1 M-step:

Note that (A.1),(A.2), and (A.3) in the complete-data log-likelihood involve separate parameters. Therefore, they can be maximized independently of one another.

Step M.1: Maximize the expected value of (A.4):

Under the constraint that $\sum_{k=1}^{K} \pi_k = 1$, we construct the Lagrangian:

$$\sum_{l=1}^{L} z_{lk} \log(\pi_k) + \lambda_\pi (\sum_{k=1}^{K} \pi_k - 1)$$

The above maximum is achieved when: $\hat{\pi}_k = \frac{1}{L} \sum_{l=1}^{L} z_{lk}$

Step M.2: Maximize the expected value of (A.3):

In this step, we can estimate μ_k, Λ_k for $k = 1, \ldots, K$ separately. For a certain k, we have:

$$\begin{aligned} B_k &\equiv E\left[\sum_{l=1}^{L} z_{lk} \left[\log|\Lambda_k| + (\gamma_l - \mu_k)^T \Lambda_k^{-1} (\gamma_l - \mu_k)\right] | Z_l = k, S_l^o\right] \\ &= \sum_{l=1}^{L} z_{lk} \log|\Lambda_k| + \sum_{l=1}^{L} z_{lk} E\left[(\gamma_l - \mu_k)^T \Lambda_k^{-1} (\gamma_l - \mu_k) | Z_l = k, S_l^o\right] \\ &= B_{k.1} + B_{k.2} \end{aligned}$$

Define $n_k = \sum_{l=1}^{L} z_{lk}$, then we have

$$B_{k.1} = n_k \log|\Lambda_k|$$

Define $\bar{\mu}_k = \sum_{l=1}^{L} [z_{lk} E(\gamma_l | Z_l = k, S_l^o)] / \sum_{l=1}^{L} z_{lk}$, then we have

$$\begin{aligned} B_{k.2} &= \sum_{l=1}^{L} z_{lk} E[(\gamma_l - \bar{\mu}_k)^T \Lambda_k^{-1} (\gamma_l - \bar{\mu}_k) + (\bar{\mu}_k - \mu_k)^T \Lambda_k^{-1} (\bar{\mu}_k - \mu_k) \\ &+ (\bar{\mu}_k - \mu_k)^T \Lambda_k^{-1} (\gamma_l - \bar{\mu}_k) + (\gamma_l - \bar{\mu}_k)^T \Lambda_k^{-1} (\bar{\mu}_k - \mu_k) | Z_l = k, S_l^o] \\ &= \sum_{l=1}^{L} z_{lk} E\left[(\gamma_l - \bar{\mu}_k)^T \Lambda_k^{-1} (\gamma_l - \bar{\mu}_k) + (\bar{\mu}_k - \mu_k)^T \Lambda_k^{-1} (\bar{\mu}_k - \mu_k) | Z_l = k, S_l^o\right] \end{aligned}$$

Therefore, $B_{k.2}$ achieves its minimum when $\hat{\mu}_k = \bar{\mu}_k$. Accordingly, we have

$$\begin{aligned}
B_{k.2} &= \sum_{l=1}^{L} E\left[z_{lk}(\gamma_l - \bar{\mu}_k)^T \Lambda_k^{-1}(\gamma_l - \bar{\mu}_k)|Z_l = k, C_l, Q_l\right] \\
&= \sum_{l=1}^{L} E\left[trace\left[z_{lk}(\gamma_l - \bar{\mu}_k)^T \Lambda_k^{-1}(\gamma_l - \bar{\mu}_k)\right]|Z_l = k, S_l^o\right] \\
&= \sum_{l=1}^{L} E\left[trace\left[\Lambda_k^{-1} z_{lk}(\gamma_l - \bar{\mu}_k)(\gamma_l - \bar{\mu}_k)^T\right]|Z_l = k, S_l^o\right] \\
&= \sum_{l=1}^{L} trace\left[\Lambda_k^{-1} z_{lk} E\left[(\gamma_l - \bar{\mu}_k)(\gamma_l - \bar{\mu}_k)^T|Z_l = k, S_l^o\right]\right] \\
&= trace\left[\Lambda_k^{-1}\left[\sum_{l=1}^{L} z_{lk} E\left[(\gamma_l - \bar{\mu}_k)(\gamma_l - \bar{\mu}_k)^T|Z_l = k, S_l^o\right]\right]\right]
\end{aligned}$$

Define $W_k = \frac{1}{n_k}\sum_{l=1}^{L} z_{lk} E\left[(\gamma_l - \bar{\mu}_k)(\gamma_l - \bar{\mu}_k)^T|Z_l = k, S_l^o\right]$, then we have $B_{k2} = n_k\left[trace(\Lambda_k^{-1} W_k)\right]$. Remember that $B_k = B_{k1} + B_{k2}$, therefore,

$$\frac{1}{n_k} B_k = \log|\Lambda_k| + trace(\Lambda_k^{-1} W_k)$$

According to the Theorem 2 in Celeux and Govaert 1993, B_k achieves its minimum if $\hat{\Lambda}_k = W_k$.

Step M.3: Maximize the expected value of (A.2):

We first set the derivative (with respective to σ^2) to 0, then we have:

$$\sum_{l=1}^{L}\sum_{k=1}^{K} z_{lk}\left(\frac{n_l}{\sigma^2} - \frac{E\left[\|S_l - B\gamma_l\|^2|(Z_l = k, S_l^o)\right]}{\sigma^4}\right) = 0$$

By solving the above equation, we have

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{\sum_{l=1}^{L}\sum_{k=1}^{K}\hat{z}_{lk} n_l}\sum_{l=1}^{L}\sum_{k=1}^{K} z_{lk} E\left[\|S_l - B\gamma_l\|^2|(Z_l = k, S_l^o)\right] \\
&= \frac{1}{\sum_{l=1}^{L} n_l}\sum_{l=1}^{L}\sum_{k=1}^{K} z_{lk} E\left[(S_l - B\gamma_l)^T(S_l - B\gamma_l)|(Z_l = k, S_l^o)\right]
\end{aligned}$$

In summary, we have:

$$\hat{\pi}_k = \frac{1}{L}\sum_{l=1}^{L} z_{lk}$$

$$\hat{\mu}_k = \bar{\mu}_k = \sum_{l=1}^{L}[z_{lk} E(\gamma_l|Z_l = k, S_l^o)]/\sum_{l=1}^{L} z_{lk}$$

$$\hat{\Lambda}_k = W_k = \frac{1}{n_k}\sum_{l=1}^{L} z_{lk} E\left[(\gamma_l - \bar{\mu}_k)(\gamma_l - \bar{\mu}_k)^T|Z_l = k, S_l^o\right]$$

$$\hat{\sigma}^2 = \frac{1}{\sum_{l=1}^{L} n_l} \sum_{l=1}^{L} \sum_{k=1}^{K} z_{lk} E\left[(S_l - B\gamma_l)^T (S_l - B\gamma_l) | (Z_l = k, S_l^o)\right]$$

All the above quantities can be computed using the result of (Step E.3).

The EM algorithm iterates through (Step E.1),(Step E.2),(Step E.3),(Step M.1),(Step M.2), (Step M.3) and (Step M.4) until all of the parameters have converged. We denote the converging values by: $\hat{\mu}_k, \hat{\Lambda}_k, \hat{\pi}_k, \hat{\sigma}^2$.

REFERENCES

[1] Araki, Y., Konishi, S., and Imoto, S., "Functional discriminant analysis for microarray gene expression data via radial basis function networks," in *Proceedings of COMPSTAT Symposium*, pp. 613–620, Citeseer, 2004.

[2] Bae, S., Kim, S., Kim, M., Lee, B., and Kang, C., "Degradation analysis of nano-contamination in plasma display panels," *Reliability, IEEE Transactions on*, vol. 57, no. 2, pp. 222–229, 2008.

[3] Bae, S., Kuo, W., and Kvam, P., "Degradation models and implied lifetime distributions," *Reliability Engineering & System Safety*, vol. 92, no. 5, pp. 601–608, 2007.

[4] Bae, S. and Kvam, P., "A change-point analysis for modeling incomplete burn-in for light displays," *IIE Transactions*, vol. 38, no. 6, pp. 489–498, 2006.

[5] Bae, S. J. and Kvam, P. H., "A nonlinear random-coefficients model for degradation testing," *Technometrics*, vol. 46, no. 4, pp. 460–469, 2004.

[6] Bagdonavicius, V. and Nikulin, M., "Estimation in degradation models with explanatory variables," *Lifetime Data Analysis*, vol. 7, no. 1, pp. 85–103, 2001.

[7] Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D., *Statistical inference under order restrictions. The theory and application of isotonic regression.* John Wiley & Sons, London-New York-Sydney, 1972. Wiley Series in Probability and Mathematical Statistics.

[8] Besse, P. and Ramsay, J. O., "Principal components analysis of sampled functions," *Psychometrika*, vol. 51, no. 2, pp. 285–311, 1986.

[9] Bian, L. and Gebraeel, N., "Stochastic methodology for prognostics under continuously varying environmental profiles," *Statistical Analysis and Data Mining*, 2012.

[10] Bigot, J., "Landmark-based registration of curves via the continuous wavelet transform," *Journal of Computational and Graphical Statistics*, vol. 15, no. 3, pp. 542–564, 2006.

[11] Bigot, J. and Gadat, S., "A deconvolution approach to estimation of a common shape in a shifted curves model," *The Annals of Statistics*, vol. 38, no. 4, pp. 2422–2464, 2010.

[12] Bogdanoff, J., Kozin, F., and Bogdanoff, J., *Probabilistic models of cumulative damage.* Wiley New York etc., 1985.

[13] Brumback, L. C. and Lindstrom, M. J., "Self modeling with flexible, random time transformations," *Biometrics*, vol. 60, no. 2, pp. 461–470, 2004.

[14] Chiou, J. and Li, P., "Functional clustering and identifying substructures of longitudinal data," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 4, pp. 679–699, 2007.

[15] Cox, D. R. and Miller, H. D., *The theory of stochastic processes*. New York: John Wiley & Sons Inc., 1965.

[16] Cross, R., Makeev, A., and Armanios, E., "A comparison of predictions from probabilistic crack growth models inferred from virklers data," *Journal of the ASTM International*, vol. 3, no. 10, 2006.

[17] Cuesta-Albertos, J. and Fraiman, R., "Impartial trimmed¡ i¿ k¡/i¿-means for functional data," *Computational statistics & data analysis*, vol. 51, no. 10, pp. 4864–4877, 2007.

[18] Davison, A. C. and Hinkley, D. V., *Bootstrap methods and their application*, vol. 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press, 1997. With 1 IBM-PC floppy disk (3.5 inch; HD).

[19] De Leeuw, J., Hornik, K., and Mair, P., "Isotone optimization in r: Pool-adjacent-violators algorithm (pava) and active set methods," 2009.

[20] Delia, M.-C. and Rafael, P.-O., "A maintenance model with failures and inspection following Markovian arrival processes and two repair modes," *European J. Oper. Res.*, vol. 186, no. 2, pp. 694–707, 2008.

[21] Di, C.-Z., Crainiceanu, C. M., Caffo, B. S., and Punjabi, N. M., "Multilevel functional principal component analysis," *Ann. Appl. Stat.*, vol. 3, no. 1, pp. 458–488, 2009.

[22] Doksum, K. and Hbyland, A., "Models for variable-stress accelerated life testing experiments based on wener processes and the inverse gaussian distribution," *Technometrics*, vol. 34, no. 1, pp. 74–82, 1992.

[23] Efron, B. and Tibshirani, R. J., *An introduction to the bootstrap*, vol. 57 of *Monographs on Statistics and Applied Probability*. New York: Chapman and Hall, 1993.

[24] Eilers, P. and Marx, B., "Flexible smoothing with b-splines and penalties," *Statistical science*, pp. 89–102, 1996.

[25] Elwany, A. and Gebraeel, N., "Real-time estimation of mean remaining life using sensor-based degradation models," *Journal of manufacturing science and engineering*, vol. 131, p. 051005, 2009.

[26] Escabias, M., Aguilera, A., and Valderrama, M., "Modeling environmental data by functional principal component logistic regression," *Environmetrics*, vol. 16, no. 1, pp. 95–107, 2005.

[27] Fan, J. and Yao, Q., *Nonlinear time series: Nonparametric and parametric methods*. Springer Verlag, 2003.

[28] Fraley, C. and Raftery, A., "Model-based clustering, discriminant analysis, and density estimation," *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 611–631, 2002.

[29] Friedman, J., "Regularized discriminant analysis," *Journal of the American statistical association*, vol. 84, no. 405, pp. 165–175, 1989.

[30] Gamboa, F., Loubes, J., and Maza, E., "Semi-parametric estimation of shifts," *Electronic Journal of Statistics*, vol. 1, pp. 616–640, 2007.

[31] Garcia-Escudero, L. and Gordaliza, A., "A proposal for robust curve clustering," *Journal of classification*, vol. 22, no. 2, pp. 185–201, 2005.

[32] Gasser, T. and Kneip, A., "Searching for structure in curve samples," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1179–1188, 1995.

[33] Gebraeel, N., "Sensory-updated residual life distributions for components with exponential degradation patterns," *Automation Science and Engineering, IEEE Transactions on*, vol. 3, no. 4, pp. 382–393, 2006.

[34] Gebraeel, N., Elwany, A., and Pan, J., "Residual life predictions in the absence of prior degradation knowledge," *Reliability, IEEE Transactions on*, vol. 58, no. 1, pp. 106–117, 2009.

[35] Gebraeel, N. and Pan, J., "Prognostic degradation models for computing and updating residual life distributions in a time-varying environment," *Reliability, IEEE Transactions on*, vol. 57, no. 4, pp. 539–550, 2008.

[36] Gebraeel, N., Lawley, M., Rong, L., and JENNIFER, K., "Residual-life distributions from component degradation signals: A bayesian approach," *IIE Transactions*, vol. 37, no. 6, pp. 543–557, 2005.

[37] Gervini, D. and Gasser, T., "Self-modelling warping functions," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 66, no. 4, pp. 959–971, 2004.

[38] Gervini, D. and Gasser, T., "Nonparametric maximum likelihood estimation of the structural mean of a sample of curves," *Biometrika*, vol. 92, no. 4, pp. 801–820, 2005.

[39] Hall, P. and Heckman, N., "Estimating and depicting the structure of a distribution of random functions," *Biometrika*, vol. 89, no. 1, pp. 145–158, 2002.

[40] Hall, P. and Hosseini-Nasab, M., "On properties of functional principal components analysis," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 68, no. 1, pp. 109–126, 2006.

[41] Horrocks, J. and Thompson, M., "Modeling event times with multiple outcomes using the wiener process with drift," *Lifetime Data Analysis*, vol. 10, no. 1, pp. 29–49, 2004.

[42] James, G. M., Hastie, T. J., and Sugar, C. A., "Principal component models for sparse functional data," *Biometrika*, vol. 87, no. 3, pp. 587–602, 2000.

[43] James, G., "Curve alignment by moments," *The Annals of Applied Statistics*, pp. 480–501, 2007.

[44] James, G. and Hastie, T., "Functional linear discriminant analysis for irregularly sampled curves," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 533–550, 2001.

[45] James, G. and Sugar, C., "Clustering for sparsely sampled functional data," *Journal of the American Statistical Association*, vol. 98, no. 462, pp. 397–408, 2003.

[46] JEFFREY, P. and STEVEN, M., "Stochastic models for degradation-based reliability," *IIE Transactions*, vol. 37, no. 6, pp. 533–542, 2005.

[47] Karhunen, K., *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*, vol. 37. Universitat Helsinki, 1947.

[48] Kass, R. and Raftery, A., "Bayes factors," *Journal of the american statistical association*, pp. 773–795, 1995.

[49] Kharoufeh, J. P., "Explicit results for wear processes in a Markovian environment," *Oper. Res. Lett.*, vol. 32, no. 2, p. 196, 2004.

[50] Kharoufeh, J. P. and Mixon, D. G., "On a Markov-modulated shock and wear process," *Naval Res. Logist.*, vol. 56, no. 6, pp. 563–576, 2009.

[51] Kharoufeh, J., Solo, C., and Ulukus, M., "Semi-markov models for degradation-based reliability," *IIE Transactions*, vol. 42, no. 8, pp. 599–612, 2010.

[52] Kneip, A., Li, X., MacGibbon, K., and Ramsay, J., "Curve registration by local regression," *Canadian Journal of Statistics*, vol. 28, no. 1, pp. 19–29, 2000.

[53] Kneip, A. and Ramsay, J., "Combining registration and fitting for functional models," *Journal of the American Statistical Association*, vol. 103, no. 483, pp. 1155–1165, 2008.

[54] Kneip, A. and Gasser, T., "Statistical tools to analyze data representing a sample of curves," *Ann. Statist.*, vol. 20, no. 3, pp. 1266–1305, 1992.

[55] Kotulski, Z., "On efficiency of identification of a stochastic crack propagation model based on virkler experimental data," *Archives of Mechanics- Archiwum Mechaniki Stosowanej*, vol. 50, no. 5, pp. 829–847, 1998.

[56] Kuniewski, S., van der Weide, J., and van Noortwijk, J., "Sampling inspection for the evaluation of time-dependent reliability of deteriorating systems under imperfect defect detection," *Reliability Engineering & System Safety*, vol. 94, no. 9, pp. 1480–1490, 2009.

[57] Lang, S. and Brezger, A., "Bayesian P-splines," *J. Comput. Graph. Statist.*, vol. 13, no. 1, pp. 183–212, 2004.

[58] Lawless, J. and Crowder, M., "Covariates and random effects in a gamma process model with application to degradation and failure," *Lifetime Data Anal.*, vol. 10, no. 3, pp. 213–227, 2004.

[59] Leng, X. and Müller, H., "Classification using functional data analysis for temporal gene expression data," *Bioinformatics*, vol. 22, no. 1, pp. 68–76, 2006.

[60] Li, G. and Luo, J., "Shock model in markovian environment," *Naval Research Logistics (NRL)*, vol. 52, no. 3, pp. 253–260, 2005.

[61] Liao, C. and Tseng, S., "Optimal design for step-stress accelerated degradation tests," *Reliability, IEEE Transactions on*, vol. 55, no. 1, pp. 59–66, 2006.

[62] Lindley, D. and Smith, A., "Bayes estimates for the linear model," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1–41, 1972.

[63] Liu, X. and Müller, H., "Modes and clustering for time-warped gene expression profile data," *Bioinformatics*, vol. 19, no. 15, pp. 1937–1944, 2003.

[64] Liu, X. and Müller, H.-G., "Functional convex averaging and synchronization for time-warped random curves," *J. Amer. Statist. Assoc.*, vol. 99, no. 467, pp. 687–699, 2004.

[65] Loader, C., *Local regression and likelihood*. Statistics and Computing, New York: Springer-Verlag, 1999.

[66] Lu, C. J. and Meeker, W. Q., "Using degradation measures to estimate a time-to-failure distribution," *Technometrics*, vol. 35, no. 2, pp. 161–174, 1993.

[67] Lu, C. and Meeker, W., "Using degradation measures to estimate a time-to-failure distribution," *Technometrics*, pp. 161–174, 1993.

[68] Lu, J., Park, J., and Yang, Q., "Statistical inference of a time-to-failure distribution derived from linear degradation data," *Technometrics*, pp. 391–400, 1997.

[69] Meeker, W., Escobar, L., and Lu, C., "Accelerated degradation tests: modeling and analysis," *Technometrics*, pp. 89–99, 1998.

[70] Montoro-Cazorla, D. and Perez-Ocon, R., "Replacement times and costs in a degrading system with several types of failure: The case of phase-type holding times," *European journal of operational research*, vol. 175, no. 2, pp. 1193–1209, 2006.

[71] Müller, H.-G. and Zhang, Y., "Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories," *Biometrics*, vol. 61, no. 4, pp. 1064–1075, 2005.

[72] Nelson, W., *Accelerated testing: statistical models, test plans and data analyses*. Wiley Online Library, 1990.

[73] Ocaña, F. A., Aguilera, A. M., and Valderrama, M. J., "Functional principal components analysis by choice of norm," *J. Multivariate Anal.*, vol. 71, no. 2, pp. 262–276, 1999.

[74] Onar, A. and Padgett, W. J., "Inverse Gaussian accelerated test models based on cumulative damage," *J. Statist. Comput. Simulation*, vol. 66, no. 3, pp. 233–247, 2000. 50th Anniversary of the Department of Statistics, Virginia Tech, Part III (Blacksburg, VA, 1999).

[75] Padgett, W. J. and Tomlinson, M. A., "Inference from accelerated degradation and failure data based on Gaussian process models," *Lifetime Data Anal.*, vol. 10, no. 2, pp. 191–206, 2004.

[76] Park, C., Koo, J., Kim, S., Sohn, I., and Lee, J., "Classification of gene functions using support vector machine for time-course gene expression data," *Computational Statistics & Data Analysis*, vol. 52, no. 5, pp. 2578–2587, 2008.

[77] Park, C. and Padgett, W., "New cumulative damage models for failure using stochastic processes as initial damage," *Reliability, IEEE Transactions on*, vol. 54, no. 3, pp. 530–540, 2005.

[78] Park, C. and Padgett, W., "Stochastic degradation models with several accelerating variables," *Reliability, IEEE Transactions on*, vol. 55, no. 2, pp. 379–390, 2006.

[79] Pettit, L. I. and Young, K. D. S., "Bayesian analysis for inverse Gaussian lifetime data with measures of degradation," *J. Statist. Comput. Simulation*, vol. 63, no. 3, pp. 217–234, 1999.

[80] Ramsay, J. O. and Silverman, B. W., *Functional data analysis.* Springer Series in Statistics, New York: Springer, second ed., 2005.

[81] Rao, C., "Some statistical methods for comparison of growth curves," *Biometrics*, vol. 14, no. 1, pp. 1–17, 1958.

[82] Ray, S. and Mallick, B., "Functional clustering by bayesian wavelet methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 2, pp. 305–332, 2006.

[83] Rice, J. A. and Silverman, B. W., "Estimating the mean and covariance structure nonparametrically when the data are curves," *J. Roy. Statist. Soc. Ser. B*, vol. 53, no. 1, pp. 233–243, 1991.

[84] Robinson, M. and Crowder, M., "Bayesian methods for a growth-curve degradation model with repeated measures," *Lifetime Data Analysis*, vol. 6, no. 4, pp. 357–374, 2000.

[85] Rønn, B., "Nonparametric maximum likelihood estimation for shifted curves," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 243–259, 2001.

[86] Rossi, F. and Villa, N., "Support vector machine for functional data classification," *Neurocomputing*, vol. 69, no. 7, pp. 730–742, 2006.

[87] Serban, N. and Wasserman, L., "Cats," *Journal of the American Statistical Association*, vol. 100, no. 471, pp. 990–999, 2005.

[88] Shiau, J. and Lin, H., "Analyzing accelerated degradation data by nonparametric regression," *Reliability, IEEE Transactions on*, vol. 48, no. 2, pp. 149–158, 1999.

[89] Silverman, B. W., "Smoothed functional principal components analysis by choice of norm," *Ann. Statist.*, vol. 24, no. 1, pp. 1–24, 1996.

[90] Singpurwalla, N., "Survival in dynamic environments," *Statistical Science*, vol. 10, no. 1, pp. 86–103, 1995.

[91] Sood, A., James, G., and Tellis, G., "Functional regression: A new model for predicting market penetration of new products," *Marketing Science*, vol. 28, no. 1, pp. 36–51, 2009.

[92] Su, C., Lu, J., Chen, D., and Hughes-Oliver, J., "A random coefficient degradation model with ramdom sample size," *Lifetime Data Analysis*, vol. 5, no. 2, pp. 173–183, 1999.

[93] Tang, R. and Müller, H., "Time-synchronized clustering of gene expression trajectories," *Biostatistics*, vol. 10, no. 1, pp. 32–45, 2009.

[94] Tang, R. and Müller, H.-G., "Pairwise curve synchronization for functional data," *Biometrika*, vol. 95, no. 4, pp. 875–889, 2008.

[95] Tarpey, T. and Kinateder, K., "Clustering functional data," *Journal of classification*, vol. 20, no. 1, pp. 93–114, 2003.

[96] Telesca, D. and Inoue, L. Y. T., "Bayesian hierarchical curve registration," *J. Amer. Statist. Assoc.*, vol. 103, no. 481, pp. 328–339, 2008.

[97] Tseng, S., Balakrishnan, N., and Tsai, C., "Optimal step-stress accelerated degradation test plan for gamma degradation processes," *Reliability, IEEE Transactions on*, vol. 58, no. 4, pp. 611–618, 2009.

[98] Tseng, S. and Peng, C., "Optimal burn-in policy by using an integrated wiener process," *IIE Transactions*, vol. 36, no. 12, pp. 1161–1170, 2004.

[99] Tseng, S. and Peng, C., "Stochastic diffusion modeling of degradation data," *Journal of Data Science*, vol. 5, no. 3, pp. 315–333, 2007.

[100] Van Noortwijk, J., "A survey of the application of gamma processes in maintenance," *Reliability Engineering & System Safety*, vol. 94, no. 1, pp. 2–21, 2009.

[101] Van Noortwijk, J., Van der Weide, J., Kallen, M., and Pandey, M., "Gamma processes and peaks-over-threshold distributions for time-dependent reliability," *Reliability Engineering & System Safety*, vol. 92, no. 12, pp. 1651–1658, 2007.

[102] Virkler, D., Hillberry, B., and Goel, P., "The statistical nature of fatigue crack propagation," tech. rep., DTIC Document, 1977.

[103] Wang, K. and Gasser, T., "Alignment of curves by dynamic time warping," *Ann. Statist.*, vol. 25, no. 3, pp. 1251–1276, 1997.

[104] Wang, K. and Gasser, T., "Synchronizing sample curves nonparametrically," *Ann. Statist.*, vol. 27, no. 2, pp. 439–460, 1999.

[105] Wang, S., Jank, W., and Shmueli, G., "Explaining and forecasting online auction prices and their dynamics using functional data analysis," *Journal of Business and Economic Statistics*, vol. 26, no. 2, pp. 144–160, 2008.

[106] Wang, W., "A model to determine the optimal critical level and the monitoring intervals in condition-based maintenance," *International Journal of Production Research*, vol. 38, no. 6, pp. 1425–1436, 2000.

[107] Wang, X., "Nonparametric estimation of the shape function in a gamma process for degradation data," *Canadian Journal of Statistics*, vol. 37, no. 1, pp. 102–118, 2009.

[108] Wang, X. and Xu, D., "An inverse Gaussian process model for degradation data," *Technometrics*, vol. 52, no. 2, pp. 188–197, 2010. With supplementary material available online.

[109] Whitmore, G., "Estimating degradation by a wiener diffusion process subject to measurement error," *Lifetime Data Analysis*, vol. 1, no. 3, pp. 307–319, 1995.

[110] Whitmore, G., Crowder, M., and Lawless, J., "Failure inference from a marker process based on a bivariate wiener model," *Lifetime Data Analysis*, vol. 4, no. 3, pp. 229–251, 1998.

[111] Whitmore, G. and Schenkelberg, F., "Modelling accelerated degradation data using wiener diffusion with a time scale transformation," *Lifetime Data Analysis*, vol. 3, no. 1, pp. 27–45, 1997.

[112] Wu, S.-J. and Shao, J., "Reliability analysis using the least squares method in nonlinear mixed-effect degradation models," *Statist. Sinica*, vol. 9, no. 3, pp. 855–877, 1999.

[113] Xu, Z., Ji, Y., and Zhou, D., "A new real-time reliability prediction method for dynamic systems based on on-line fault prediction," *Reliability, IEEE Transactions on*, vol. 58, no. 3, pp. 523–538, 2009.

[114] Yang, K. and Yang, G., "Degradation reliability assessment using severe critical values," *INTERNATIONAL JOURNAL OF RELIABILITY QUALITY AND SAFETY ENGINEERING*, vol. 5, pp. 85–95, 1998.

[115] Yao, F. and Lee, T. C. M., "Penalized spline models for functional principal component analysis," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 68, no. 1, pp. 3–25, 2006.

[116] Yao, F., Müller, H.-G., Clifford, A. J., Dueker, S. R., Follett, J., Lin, Y., Buchholz, B. A., and Vogel, J. S., "Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate," *Biometrics*, vol. 59, no. 3, pp. 676–685, 2003.

[117] Yao, F., Müller, H.-G., and Wang, J.-L., "Functional data analysis for sparse longitudinal data," *J. Amer. Statist. Assoc.*, vol. 100, no. 470, pp. 577–590, 2005.

[118] Yeung, K., Medvedovic, M., Bumgarner, R., and others, "Clustering gene-expression data with repeated measurements," *Genome Biology*, vol. 4, no. 5, p. R34, 2003.

[119] Zhou, R., Gebraeel, N., and Serban, N., "Degradation modeling and monitoring of truncated degradation signals," *IIE Transactions*, vol. 44, no. 9, pp. 793–803, 2012.

[120] Zhou, R. R., Serban, N., and Gebraeel, N., "Degradation modeling applied to residual lifetime prediction using functional data analysis," *Ann. Appl. Stat.*, vol. 5, no. 2B, pp. 1586–1610, 2011.